MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0361316 9

54

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION
GKE

ORNL-2704
Physics and Mathematics

A GLOSSARY FOR NUMERICAL ANALYSIS

Alston S. Householder

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this document, send in name with document and the library will arrange a loan.

OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission to the extent that such employee or contractor prepares, handles or distributes, or provides access to, any information pursuant to his employment or contract with the Commission.

ADDITIONAL ERRATA

P. 8, line 7: latter portion of equation should read "... = $P^{(m-1)}(r) \neq P^{(m)}(r)$,".

P. 10, lines 3, 4, and 5 from below: close up spaces between δa and δb where spaces occur.

P. 11, line 2: "given" should be "Given."

P. 12, line 7: after "Euler-Maclaurin formula", add "(NO)." deleting period after "formula."
" , line 18: preceding the entry "Bessel interpolation formula" and as a separate line insert the words "See NO."
"$E^u \approx E^{\frac{1}{2}}$" should read "$E^u = E^{\frac{1}{2}}$".

P. 13, line 9: after "in the form" insert "(see <u>star</u>)".

P. 14, line 11: "given" should read "Given."
" , line 15: "Descarte's" should read "Descartes's."
" , line 21: "in" should read "In."

P. 16, line 12: equation should read "$\int_a^b w(x) f(x)\, dx = \omega \sum_0^n f(x_i) + R_n$,".

P. 17, line 4: close spaces between "nx."
" , line 2 from below: "for" should read "For."

P. 18, line 7: "in" should read "In."

P. 19, line 7: insert the fourth paren - it should read "$(\|\delta A\| / \|A\|)$".
" , line 15: delete the words "and distances."
" , lines 18 and 19: delete the sentence "Preservation of distances means that the arc length from P' to P_1' is the same as that from P to P_1."

P. 20, line 2: change "converge" to "is exact."
" , line 11: "socalled" should read "so called."
" , line 6 from below: "consider" should read "Consider."

P. 22, line 6: "a" should read "A."

P. 23, line 3: "suppose" should read "Suppose."

Corrected June 26, 1961. NB

P. 24, line 5: after "smoothing" add ", Ni."
" , line 14: "let" should read "Let."
" , line 16: "k" should read "h."
" , line 17: equation should read "$\phi_h = a_0 + a_1 h^2 + a_2 h^4 + \ldots$".
" , line 20: equation should read "$\phi_{h_1, h_2} \equiv$ - - - - - ".

P. 25, line 19: "Descarte's" should read "Descartes's."

P. 26, line 15: "for" should read "For."

P. 29, line 19: "Sterling" should read "Stirling."

P. 30, line 5: "for" should read "For."

P. 31, line 20: After sentence ending "proper value," add "But usage varies even for integral equations."

P. 32, line 7: replace "Poisson" with "Laplace."

P. 33, after line 8: insert the definition "<u>Euclidean algorithm</u> -- see <u>highest</u> <u>common</u> <u>factor</u>."
" , line 11: "= 0.577215665..." should read "≐ 0.577215665".
" , line 12: after "<u>Euler-Maclaurin formula</u> --" add "the formal expansion".

P. 34, line 2 from below: "Sterling" should read "Stirling."

P. 35, line 11: "a series of the form" should read "A general series of the form."
" , line 12: after the equation, insert (on the next line) the words "is called a trigonometric series."
" , line 20: change the period at the end of the equation to a comma and add (on the next line) "and with a_p and b_p so defined the series is called a Fourier series."
" , last equation at bottom of page: insert a dot over the equals sign and insert "n" above the "summation" and "0" below it.

P. 36, line 2: insert "n" over the "summation" and "j = 0" below it.
" , line 14: replace "weight" with "density."

P. 37, lines 6 and 7: "Siedel" should read "Seidel."

P. 39, line 1: "<u>Green function</u>" should read "<u>Green's function</u>."
" , line 4: at the end of the equation, replace the period with a comma and add (next line) "the integration extending over the interval on which the solution is to be defined."
" , line 15: after "160" add "+ ..." and delete the next line of "dots."

P. 40, line 2: insert "n" over the "summation" and "0" below it.
" , line 5: after "formula." add the sentence "More generally, higher derivatives may also be prescribed."

P. 41, line 7: after "However," insert "if every $\beta_i \neq 0$."

P. 44, line 10: insert the right paren to read "$= \beta_1 (\phi, \psi_1) + ---$".
" , line 13: "determination" should read "approximation."
" , line 21: after "difference operators replace comma with semicolon.

P. 45, line 1: "determination" should read "approximation."
" , line 21: should read "$|\phi'(x)| \leq k < 1$".

P. 46, line 21: over the large "pi" insert "0, n".

P. 47, line 2: replace "weight" with "density."
" , line 12: after "0 0" add "...".

P. 49, line 5: following the equation, on the next line insert "the summation extending over all observations,".
" , line 16: replace "weight" with "density."
" , line 17, equation should read "$(n+1) P_{n+1}(x) - (2n+1) x P_n(x) + n P_{n-1}(x) = 0$,".

P. 51, line 4 from below: replace "continuous" with "bounded."

see p. 65
P. 64, line 1: "The" should read "the."

P. 65, line 13: "d(y)" should read "F(y)".
" , line 16: delete the comma at the end of the equation and add "$+ R_n$,".
" , line 22: "Sterling" should be "Stirling."

P. 68, line 15: "a problem" should read "A problem."

P. 70, line 13: replace the words "possesses partial derivatives of all orders" with "is analytic in x and in y."
" , line 16: replace "known" with "given" and after "itself" add "one obtains."

P. 71, line 18: "given" should read "Given."

P. 72, line 18: replace "weight" with "density."

P. 73, line 12: "for" should read "For."

P. 75, following line 10: add a new sentence as follows: "The coefficients a, b, and c are functions of x and y, and the inequalities must hold throughout the region under consideration."

P. 76, lines 9 and 10: replace "in the following way:" with "by the following stars:".

P. 78, line 13: "for" should read "For."

P. 79, line 21: in equation (1) "b_n" should read "b_m".

P. 84, line 13: insert a comma after "conceal them."

P. 85, line 14: "three-eights" should read "three-eighths."

P. 88, line 1: "colum" should be "column."

P. 92, line 5: entire line should read "Analogous simpler schemes, some of lower and some of higher accuracy, are available (see Co). In this."

P. 93, line 5: insert a vertical bar "|" in front of "y*" instead of the paren.

derivatives by differences used in the numerical solution of partial differential equations.

P. 99, line 8: underline "predictor."

" , after line 12: insert a new entry as follows:

Sturm sequence -- a sequence of polynomials associated with a given polynomial possessing the properties required for applying the Sturm theorem, that no consecutive pair vanish simultaneously, and that when one vanishes the adjacent ones are opposite in sign.

" , line 14: underline "Euclidean algorithm."

" , line 15: "In the sequence" should read "In the Sturm sequence."

P. 100, line 1: "in" should read "In."

P. 101, line 11: "to" should be "To."

P. 102, line 2: raise the left paren so that it reads "(n+1)."
" , line 6: "$+\sum\sum$" should read "$+\frac{1}{2!}\sum\sum$"

P. 103, line 3: after "$\leq$ b." add "h = b - a." replacing the period after "b" with a comma.
" , line 4: replace "null" with "zero."
" , line 6: change "when also all diagonal elements are units." to read "when also every diagonal element = 1."
" , line 11: "formulas" should be "Formulas."
" , line 12: change the colon after "obtained" to a period.
" , line 16: "$j \neq k$" should read "$j \neq i$".
" , following line 18: add a new entry as follows:

<u>Trigonometric series</u> -- see <u>Fourier series</u>.

P. 107, last line: "34d" should read "3rd."
" , add two new references:

Kaj Leo Nielsen (1956): Methods in numerical analysis, New York, Macmillan, 328 pp. - Ni

Niels Erik Nörlund (1924): Vorlesungen über Differenzenrechnung, Berlin, Springer, ix + 551 pp. - NO

August 12, 1959

ORNL-2704

A Glossary for Numerical Analysis

ERRATA

Delete hyphens and close throughout in "non-null," "non-negative," "non-hermitian," "non-normal."

P. 30, line 3 - For "or scalars" read "on scalars."

" , line 4 from below - For "thus" read "then."

P. 34, line 7 from below - For "both must be taken" read "must be taken both."

P. 39, line 6 from below - For "total step" read "total-step."

P. 59, second matrix - For "L_{22}" read "I." Just below, delete "L_{11} and" and "both."

P. 65, line 4 from below - Underline "trapezoidal rule, Simpson's rule."

P. 92, line 5 - For "Analagous" read "Analogous."

P. 93, line 5 - Replace first paren by vertical bar.

P. 97, entry "Stirling formula" - Replace "ln" by "log" throughout.

P. 102, line 6 - Insert factor "$\frac{1}{2!}$".

P. 106, line 7 - For "Konvergente" read "konvergente."

" , reference to Hestenes and Stiefel paper should be dated 1952.

March 17, 1959

ORNL-2704

Contract No. W-7405-eng-26

MATHEMATICS PANEL

A GLOSSARY FOR NUMERICAL ANALYSIS

Alston S. Householder

DATE ISSUED

MAR 19 1959

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0361316 9

A Glossary for Numerical Analysis

FOREWORD

Probably no two of those who work in a field they consider "numerical analysis" will agree as to the boundaries of this field, hence as to the terms that should be included in or left out of a glossary of terms in the field. Thus a book published in 1955 and entitled simply "Numerical Analysis" contains no chapter on the solution of algebraic equations, but only a few pages of "Notes" in an appendix. A Russian book published in 1953 entitled "Numerical Methods of Mathematical Analysis" makes no reference to the subject at all. Lanczos, who does consider the subject, entitles his book "Practical Analysis," and proposes a term "parexic analysis" to apply to an area intermediate between classical mathematical analysis and what he would consider to be strict "numerical analysis." The point of view adopted here in the preparation of this glossary is not far from that urged by Lanczos, although he would doubtless consider many of the terms included here to belong more properly to parexic analysis than to numerical analysis. A further distinction must be made between numerical analysis and programming theory, an area that was hardly recognized to exist prior to the decade that is soon to end.

A few terms have been included, but only a few, that belong more properly to programming theory than to numerical analysis. There the lines are more easily drawn. In the other direction, most of what is to be found in Collatz's "Numerische Behandlung" is considered to belong more to parexic than to numerical analysis.

And since "numerical analysis" is assumed to refer to the application of arithmetic processes, there is little here on the use of analog devices, including nomograms.

By the same rule, however, it is assumed that the first step in preparing a problem for digital computation is to replace continuous variables, where they occur, by discrete variables. Hence the emphasis is upon the treatment of discrete variables (vectors in a finite space, matrices of finite order), and upon the more elementary aspects of approximation theory. Anything else has been considered either out of bounds for a mere glossary, or belonging at least to parexic, if not to pure, analysis.

Nevertheless, even within the area thus circumscribed it may be that many important terms have been overlooked. Hence in view of a possible supplement, suggestions, criticisms, and proposed additions are hopefully solicited.

Accelerating convergence -- of series and sequences. See summation of series.

Accuracy -- the degree of exactness actually possessed by an approximation, measurement, etc. Contrasted with precision, which is the degree of exactness with which the quantity is expressed. E.g. as a value of π, the number 3.1428 is more precise than accurate.

Adams-Bashford method -- for solving an ordinary differential equation, $y' = f(x, y)$, expressed in terms of difference operators, uses the predictor

$$y_{\nu+1} = y_\nu + h[f_\nu + \tfrac{1}{2} \nabla f_\nu + \tfrac{5}{12} \nabla^2 f_\nu + \tfrac{3}{8} \nabla^3 f_\nu + \ldots],$$

and the corrector

$$y_{\nu+1} = y_\nu + h[f_{\nu+1} - \tfrac{1}{2} \nabla f_{\nu+1} - \tfrac{1}{12} \nabla^2 f_{\nu+1} - \tfrac{1}{24} \nabla^3 f_{\nu+1} - \ldots].$$

The first amounts to approximating

$$\int_{x_\nu}^{x_{\nu+1}} f(x, y)\, dx$$

by means of a quadrature formula which utilizes values of f at x_ν and to the left only, the second utilizes also the value of f at $x_{\nu+1}$. The corrector is to be used with the best current approximation of $y_{\nu+1}$, and reapplied until the next approximation does not differ significantly from the previous. The formulas have significance when y and f are vectors (see Mi, Co).

Adams-Störmer method -- see Störmer method.

Address -- the symbol designating a particular cell in the storage unit of a stored-program computer.

Aitken δ^2-process -- see Delta-square process.

<u>Aitken method of interpolation</u> -- is to form a table

$$
\begin{array}{lllll}
x_0 - x & y_x(x_0) & & & \\
x_1 - x & y_x(x_1) & y_x(x_0, x_1) & & \\
x_2 - x & y_x(x_2) & y_x(x_0, x_2) & y_x(x_0, x_1, x_2) & \\
x_3 - x & y_x(x_3) & y_x(x_0, x_3) & y_x(x_0, x_1, x_3) & y_x(x_0, x_1, x_2, x_3) \\
 & \cdot & \cdot & \cdot & \cdot \quad \cdot \quad \cdot \quad \cdot
\end{array}
$$

by means of the relations $y_x(x_i) = y_i$ and

$$y_x(x_i, x_j) = \begin{vmatrix} y_x(x_i) & x_i - x \\ y_x(x_j) & x_j - x \end{vmatrix} \Big/ (x_j - x_i),$$

$$y_x(x_i, x_j, x_k) = \begin{vmatrix} y_x(x_i\ x_j) & x_j - x \\ y_x(x_i, x_k) & x_k - x \end{vmatrix} \Big/ (x_k - x_j),$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

Then $y_x(x_i, x_j)$ is the result of linear interpolation based upon the abscissae x_i and x_j; $y_x(x_i, x_j, x_k)$ that of quadratic interpolation based upon the abscissae x_i, x_j, and x_k; .... The process may be terminated whenever consecutive entries in a column agree to sufficiently many figures.

It is not necessary that the abscissae be equally spaced, or even ordered, but only that they be distinct. The method of Neville is based upon the same principle but utilizes the table

$$
\begin{array}{llllll}
x_0 - x & y_x(x_0) & & & & \\
 & & y_x(x_0, x_1) & & & \\
x_1 - x & y_x(x_1) & & y_x(x_0, x_1, x_2) & & \\
 & & y_x(x_1, x_2) & & & y_x(x_0, x_1, x_2, x_3) \; . \\
x_2 - x & y_x(x_2) & & y_x(x_1, x_2, x_3) & & \\
 & & y_x(x_2, x_3) & & & \\
x_3 - x & y_x(x_3) & & & &
\end{array}
$$

<u>Algebraic equations</u> -- an equation, or set of <u>simultaneous equations</u>, in which the unknowns occur as rational functions only. Hence the equations are expressible by equating polynomials to zero. Here the case of a single equation in one unknown will be considered:

$$a_0 \, P(x) \equiv a_0 \, x^n + a_1 \, x^{n-1} + \ldots + a_n = 0, \qquad a_0 \neq 0,$$

where the a_i do not depend upon x, and are called the coefficients of the equation, while n is the degree. This equation is equivalent to

$$P(x) \equiv x^n - c_1 \, x^{n-1} + \ldots + (-1)^n \, c_n = 0,$$

$$c_i = (-1)^i \, a_i \, / \, a_0,$$

where the c_i are the elementary <u>symmetric functions</u> of the roots x_i.

The remainder theorem states that if $a_0 \, P(x)$ is divided by $x - r$, the remainder is $a_0 \, P(r)$:

$$a_0 \, P(x) \equiv (x - r) \, Q(x) + a_0 \, P(r);$$

the factor theorem is a corollary and states that if r is a root of $P(x) = 0$, then $x - r$ divides $P(x)$. The "fundamental theorem of algebra" states that every algebraic equation has a root, real or complex. These theorems imply that $P(x)$ can be factored completely:

$$P(x) = (x - x_1) \, (x - x_2) \, \ldots \, (x - x_n),$$

each x_i being a root. If $x_i = x_j$ for some $i \neq j$, then $x_i = x_j$ is a double root; if $x_i = x_j = x_k$, a triple root, ... Counting a root x_i of multiplicity m as being m coincident roots, one says that an algebraic equation of degree n has exactly n roots (neither more nor less). A <u>Taylor series</u> expansion gives

$$P(x - r) = P(r) + (x - r)\, P'(r) + \ldots + (x - r)^n\, P^{(n)}(r) / n!$$

Hence r is a root of multiplicity m if and only if

$$0 = P(r) = P'(r) = \text{"} \ldots = P^{(m-1)}(r) \neq P^{(m)}(r), \text{"}.$$

hence r satisfies the derived equations

$$P^{(i)}(x) = 0, \qquad i = 0, 1, \ldots, m - 1.$$

In general, setting $y = x - r$, if

$$b_i = P^{(n-i)}(r) / (n - i)!\ ,$$

then any root of

$$y^n + b_1\, y^{n-1} + \ldots + b_n = 0$$

is r less than a root x_i, hence the roots are said to have been reduced by r. Repeated <u>synthetic division</u> can be applied to the original equation to obtain the b_i, since $b_n = P(r)$ is the remainder after dividing $P(x)$ by $x - r$; b_{n-1} that after dividing the quotient by $x - r$; ....

Other useful transformations are the following:

The roots of

$$a_0\, y^n - a_1\, y^{n-1} + \ldots + (-1)^n\, a_n = 0$$

are the negatives of the roots of the original; those of

$$a_0\, y^n + a_1\, \alpha\, y^{n-1} + \ldots + a_n\, \alpha^n = 0$$

are α times the roots of the original; those of

$$a_n\, y^n + a_{n-1}\, y^{n-1} + \ldots + a_0 = 0$$

are the reciprocals of those of the original. These can be derived by setting $x = -y$, $x = y/\alpha$, and $x = 1/y$.

When the coefficients a_i are integers, then any rational root when expressed as a fraction p/q in lowest terms is such that p divides a_n and q divides a_0 without remainder. In particular, any integral root must be a divisor of a_n. In principle all rational roots can be found exactly, and once any root r is known, all other roots must satisfy $Q(x) = 0$ where $Q = P / (x - r)$. For irrational roots see Budan theorem, Sturm theorem, Horner method, Newton method, Bernoulli method, Graeffe method, iterative methods.

While these methods, except for Horner's, apply as well to complex roots as to real, it may be convenient to evaluate P(z) with $z = x + i\,y$, and write

$$P(z) = R(x, y) + i\,y\,J(x, y),$$

after collecting real and pure imaginary terms. Then

$$R(x, y) = J(x, y) = 0$$

are two simultaneous equations in x and y, and any real solution (x, y) determines a complex solution $z = x + i\,y$ of $P(z) = 0$. More to be recommended, however, is the Bairstow method for complex roots, which is, in fact, an adaptation of Newton's method to complex roots.

Algorithm -- a corruption of algorism, which meant originally the art of computing with Arabic numerals. Algorithm has now come to apply to any rule of computation, whether algebraic or numerical.

Analog computer -- a computing device in which quantities are directly represented by physical magnitudes: distances, voltages, currents, rotations, etc. Hence the mathematical operations are direct operations upon these magnitudes. Examples are slide rules, differential analyzers, and most special purpose computers, as well as wind tunnels and model basins. Analog computers are to be contrasted with digital computers, although hybrid forms exist. Most often, as in digital dif-

ferential analyzers, in overall use the machine will appear as an analog machine but will contain components that operate digitally.

Analog-to-digital converter -- any device for providing a digital representation of quantities initially represented in analog form, i.e., by physical magnitudes.

Arithmetic unit -- the portion of a stored-program computer devoted to the actual execution of the commands.

Asymptotic series -- a generally divergent series of the form

$$a_0 + a_1 / z + a_2 / z^2 + \ldots$$

which, however, represents a function of (z) in the sense that

$$\lim_{|z| \to \infty} \left[z^n \left(f(z) - \sum_0^n a_i / z^i \right) \right] = 0.$$

It generally happens that for sufficiently large z a limited number of terms of the series will yield a close approximation to f(z) whereas more terms will not.

Backward differences -- see difference operators.

Bairstow method -- for finding complex roots of an algebraic equation. Let $z^2 + a z + b$ be a trial divisor of f(z) and form

$f(z) = (z^2 + a z + b)^2 Q(z) + (z^2 + a z + b) q(z) + r(z)$, where

$$r(z) = r_1 z + r_0, \qquad q(z) = q_1 z + q_0.$$

This means that r is the remainder after dividing f by $z^2 + a z + b$, and q the remainder after dividing the quotient. Solve

$$(a q_1 - q_0) \delta a - q_1 \delta b = -r_1,$$

$$b q_1 \delta a - q_0 \delta b = -r_0$$

for δa and δb; then $z^2 + (a + \delta a) z + b + \delta b$ will be, in general, closer to a true divisor. The method is an adaptation of the Newton method to finding complex roots and was originally described by Bairstow; later by Hitchcock.

Base -- see radix.

Bernoulli method -- Given the algebraic equation

(1) $$x^n + a_1 x^{n-1} + \dots + a_n = 0,$$

let $h_0, h_1, \dots, h_{n-1}$ be arbitrary numbers, not all zero, and form $h_n, h_{n+1}, \dots,$ by

$$h_{n+\nu} + a_1 h_{n+\nu-1} + \dots + a_n h_\nu = 0.$$

If (1) has a unique root of largest modulus, then in general the quotients $h_{\mu+1} / h_\mu$ approach that root. The method can be extended to transcendental equations. Let

(2) $$f(z) \equiv 1 + c_1 z + c_2 z^2 + \dots$$

converge in some circle about the origin in the complex plane, and let

(3) $$g(z) \equiv g_0 + g_1 z + g_2 z^2 + \dots$$

represent any function analytic in the same circle, and having no zero in common with $f(z)$. Let

$$\begin{aligned} h_0 &= g_0, \\ c_1 h_0 + h_1 &= g_1, \\ c_2 h_0 + c_1 h_1 + h_2 &= g_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{aligned}$$

Then if $f(z)$ has a unique zero of smallest modulus lying within that circle, then $h_\mu / h_{\mu+1}$ approaches that zero. If there are two zeros whose moduli are less than those of all others, then the roots of

$$\begin{vmatrix} z^2 & h_\mu & h_{\mu+1} \\ z & h_{\mu+1} & h_{\mu+2} \\ 1 & h_{\mu+2} & h_{\mu+3} \end{vmatrix} = 0$$

approach those zeros of $f(z)$. Likewise one can form cubics whose roots approach the three smallest roots, ... (see Ho). The Q D algorithm improves upon this principle (NBS).

Bernoulli numbers -- coefficients B_ν of the Bernoulli polynomials. Except for B_1, those with odd subscripts vanish. The values of the first several are

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30},$$

$$B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30}, \quad B_{10} = \frac{5}{66}.$$

They are the coefficients in the expansion

$$u / (e^u - 1) = \sum_0^\infty B_\nu u^\nu / \nu ! .$$

They occur in the Euler-Maclaurin formula (NO).

Bernoulli polynomials -- $B_\nu(x)$ of degree ν occur as coefficients in the expansion

$$u e^{xu} / (e^u - 1) = \sum_0^\infty u^\nu B_\nu(x) / \nu ! .$$

They satisfy the differential equations with boundary conditions

$$B'_\nu(x) = \nu B_{\nu-1}(x), \qquad B_\nu(1) = (-1)^\nu B_\nu(0),$$

and can be written symbolically in the form $B_\nu(x) = (B + x)_\nu$, where the right member is to be expanded as though ν were an exponent:

$$(B + x)_\nu = B_\nu + \nu x B_{\nu-1} + \frac{\nu(\nu-1)}{2} x^2 B_{\nu-2} + \dots$$

and the coefficients are the Bernoulli numbers. In the same fashion a recursion for the Bernoulli numbers is

$$B_\nu = (B + 1)_\nu, \qquad \nu > 1.$$

SEE NO.

Bessel interpolation formula -- $E^u = E^{\frac{1}{2}} \Big[\mu + (u - \tfrac{1}{2})\,\delta + \binom{u}{2} \mu\, \delta^2$

$$+ \frac{u - \frac{1}{2}}{3} \binom{u}{2} \delta^3 + \binom{u+1}{4} \mu\, \delta^4 + \frac{u - \frac{1}{2}}{5} \binom{u+1}{4} \delta^5 + \dots \Big]$$

(for notation see difference operators).

The formula is most useful for $0 < u < 1$. Taking $v = u - \frac{1}{2}$, an alternative, and more symmetric, expression is

$$E^u = E^{\frac{1}{2}} \Big[\mu + v\,\delta + \tfrac{1}{2!}\,(v^2 - \tfrac{1}{4})\,\mu\,\delta^2 + \tfrac{1}{3!}\,v(v^2 - \tfrac{1}{4})\,\delta^3 + \tfrac{1}{4!}\,(v^2 - \tfrac{1}{4})\,(v^2 - \tfrac{9}{4})\,\mu\,\delta^4 + \tfrac{1}{5!}\,v(v^2 - \tfrac{1}{4})\,(v^2 - \tfrac{9}{4})\,\delta^5 + \ldots \Big] .$$

Bigit -- a binary digit. This term has been proposed in order to reserve bit for use only as a measure of information.

Biharmonic equation -- the partial differential equation

$$\partial^4 \varphi / \partial x^4 + 2\,\partial^4 \varphi / \partial x^2\,\partial y^2 + \partial^4 \varphi / \partial y^4 = 0$$

satisfied by small deflections of a plate. The usual coefficients for a square mesh are in the form (see star).

$$\begin{array}{ccccc} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & 20 & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{array}$$

Binary -- characterized by two alternatives; compounded of two parts. In the binary representation of numbers, the radix is 2, this being used in many digital computers.

Binary point -- the radix point in the representation of a number with the radix 2.

Biquinary -- a form of number representation combining the radices 2 and 5.

Bit -- a unit of measure of information; viz., that amount represented by one of two alternatives. A binary digit, i.e., 0 or 1, in number representation with the radix 2.

Block relaxation -- a variant of the method of relaxation for solving linear systems in which one solves two or more equations of the set simultaneously for the corresponding unknowns in terms of current approximations to the others. This generally speeds convergence if convergence occurs.

Blunder -- any malfunction, human or otherwise, such as faulty arithmetic, use of an incorrect formula, etc., in a computation. Also the result thereof. See error.

Boundary value problem -- that of finding the (or a) function satisfying an ordinary or a partial differential equation and satisfying also a condition or conditions at the boundaries. Elliptic differential equations are ordinarily of this form. For ordinary differential equations the usual method of solution is to approximate the derivative by finite differences and solve the resulting difference equations. This leads to matrix inversion and the solution of linear equations, and possibly to the need for finding proper values and vectors.

Budan theorem -- Given any polynomial $P(x)$ in the single variable x, let P and its derivatives be evaluated at $x = r$, for any r. Ignoring vanishing derivatives, let V_r represent the number of variations in sign in the sequence $P, P', P'', \ldots$. Then the number of roots of $P(x) = 0$ exceeding r is $V_r + 2\nu$, where ν is an integer ≥ 0. Descartes's rule of signs is a particular case with $r = 0$, where V_0 is the same as the variations in sign in the coefficients. Any r for which $V_r = 0$, all $P^{(i)}$ having the same sign, is an upper bound for the roots of $P(x) = 0$. In case $V_r - V_s = 1$, there must be exactly one root on the interval from r to s. All roots are counted according to multiplicity, e.g., a double root is counted twice.

Calculator -- In current usage the term is used interchangeably with computer as either one who or that which calculates or computes. Since the term "computer" more often designates a machine in current literature, it is suggested here that "calculator" be applied only to the human being, and "computer" only to the machine.

Carry -- in digital addition of two numbers, an increase by one of the sum of the digits in any position when the sum in the position to the right exceeds the base.

Čebyšev -- see Chebyshev.

Central difference operator -- see difference operators.

Central mean operator -- see difference operators.

Characteristic polynomial -- of a matrix, see matrix.

Characteristic value -- of a matrix, see matrix, proper values and vectors, and principal vector.

Characteristic vector -- of a matrix, see matrix, principal vector, proper values and vectors.

Chebyshev approximation -- see minimax approximation.

Chebyshev expansion -- an expansion in Chebyshev polynomials. Since $T_\nu = x^\nu +$ terms of lower degree, therefore $x^\nu = T_\nu +$ Chebyshev polynomials of lower degree. Let $f(x)$ have the expansion

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots, \qquad -1 \le x \le 1.$$

By expressing each x^ν in terms of Chebyshev polynomials and collecting terms one has

$$f(x) = \beta_0 + \beta_1 T_1(x) + \beta_2 T_2(x) + \ldots.$$

It generally happens that for a given degree of accuracy this Chebyshev expansion can be truncated much sooner than the power series expansion, hence for computing purposes is more effective. See minimax approximation (for more details see Ko, La).

Chebyshev polynomials -- see also Chebyshev system, minimax approximation.

$$T_0 = 1$$

$$T_\nu(x) = 2^{1-\nu} \cos \nu\, \theta, \qquad x = \cos \theta.$$

Hence the polynomial T_ν is of degree ν, with leading coefficient 1, and

$$|T_\nu(x)| \le 2^{1-\nu} \quad \text{for} \quad -1 \le x \le 1.$$

They satisfy the recursion

$$4\ T_{\nu+1} = 4\ x\ T_\nu - T_{\nu-1},$$

and T_ν, further, possesses the important minimax property that among all polynomials of degree ν and leading coefficient 1, the maximum absolute value of T_ν on the interval from -1 to +1 is least.

One should note that in some books the polynomials

$$P_\nu(x) = 2^{\nu-1}\ T_\nu(x) = \cos\ n\ \theta$$

are called the Chebyshev polynomials. For these the recursion is

$$P_{\nu+1} = 2x\ P_\nu - P_{\nu-1}$$

(see Ko, La).

<u>Chebyshev quadrature formula</u> -- a formula of the form

$$\int_a^b w(x)\ f(x)\ dx \doteq \omega \sum_0^n f(x_i) + R_n\ ,$$

where the points x_i are so chosen that the integration is exact when $f(x)$ is a polynomial of degree as high as possible. It turns out the x_i can be determined as roots of an algebraic equation of degree $n + 1$, but unless $n \leq 7$ or $n = 9$ they do not all lie on the interval of integration. The formula is advantageous in particular in cases where the $f(x_i)$ are experimentally measured, since in the formula they are equally weighted.

<u>Chebyshev system</u> -- of order n, a set of functions

$$\varphi_0(x),\ \ \varphi_1(x),\ \ldots,\ \varphi_n(x),$$

linearly independent and continuous on an interval (a, b) with the property that if the linear combination

$$F_n(x) = \alpha_0\ \varphi_0(x) + \ldots + \alpha_n\ \varphi_n(x)$$

vanishes more than n times on the interval, then it vanishes identically, and

hence $\alpha_0 = \alpha_1 = \ldots = \alpha_n = 0$. Such a system is

$$\varphi_0 = 1, \quad \varphi_1 = x, \ \ldots, \quad \varphi_n = x^n ;$$

for the interval from 0 to 2π, the system

$$\varphi_0 = 1, \quad \varphi_1 = \sin x, \quad \varphi_2 = \cos x, \ \ldots, \quad \varphi_{2n-1} = \sin nx, \quad \varphi_{2n} = \cos nx$$

is a Chebyshev system of order 2n. Such systems are of importance in minimax approximation.

Chió's method -- for evaluating a determinant is identical in principle with Gaussian elimination for matrix inversion and the solution of linear equations. If the elements are α_{ij} and $\alpha_{11} \neq 0$, then from the ith row subtract the first multiplied by $\alpha_{i1} / \alpha_{11}$. The result is a determinant of equal value with a zero in position (i, 1). After this has been done for each row other than the first, the determinant reduces to one of lower order.

Choleski's method -- for inverting a positive definite Hermitian matrix A is to express it first in the form

$$A = R^* R$$

where R is upper triangular and R* its conjugate transpose, and thence to compute R^{-1} and multiply by its conjugate transpose. The method is mathematically equivalent to the method of elimination, but requires the extraction of n square roots.

Closed quadrature formula -- one which explicitly includes the values of the ordinates at the endpoints, to be contrasted with an open formula in which the ordinate of at least one endpoint does not appear. Closed formulas are used as correctors, open formulas as predictors, in the numerical solution of ordinary differential equations.

Code -- For any given stored-program computer, each operation of its repertory is represented by a code symbol, which is a particular sequence of characters. Also

each cell in the storage unit is designated by a code symbol, the address of the cell. An ordered sequence of symbols representing operations and operands (in the form of addresses) when represented in the storage unit, directs the computer to perform the operations upon the operands in the specified sequence, and this ordered set of code symbols is a machine code, and it defines a routine (or program).

Complement -- In decimal representation the complement of a number x is $10^{\nu} - x$, where ν is some fixed integer, positive, negative, or zero. In binary representation it is $2^{\mu} - x$, where μ is an integer. The subtraction of a number is thus essentially equivalent to the addition of its complement, and in computer construction it is generally easier to mechanize the formation of a complement than that of an arbitrary difference. In binary notation, the one's-complement of x is the number obtained when each digit x_i of x is replaced by $1 - x_i$.

Computer -- see calculator. Any physical device used as an aid in computing. It is customary to distinguish two general classes, analog computers and digital computers.

Condition number -- an indicator of numerical stability. In particular, the condition number of a matrix is usually taken to be a number that is equal to unity for a scalar matrix, and that is large when the matrix is in some sense nearly singular. Hence if the problem is to solve the linear system

$$A\,x = h,$$

the problem is unstable in case the condition number of A is large. Among the several functions of the matrix that have been proposed, the most suitable seems to be

$$\gamma(A) = \|A\|\ \|A^{-1}\|$$

for some appropriately chosen norm. The reason for this is that for small changes

in A, the ratio of the relative change in the solution to the relative change in A can be shown not to exceed $\gamma(A)$. This is to say that if δ A represents a small change in A, and δ x the induced change in x, then

$$(\|\delta x\| / \|x\|) / (\|\delta A\| / \|A\|) \leq \|A\| \, \|A^{-1}\| = \gamma(A),$$

provided second-order effects are neglected. More exactly, if $E = A^{-1} \delta A$ and $\|E\| < 1$, then

$$(\|\delta x\| / \|x\|) / (\|\delta A\| / \|A\|) \leq \gamma(A) / (1 - \|E\|).$$

It should be observed that this condition number relates only to the invertibility of a matrix, and has no relation to the finding of its <u>proper values and vectors</u>. If A is diagonalizable,

$$A = P \Lambda P^{-1}$$

where Λ is diagonal, then $\|P\| \, \|P^{-1}\|$ seems to be the appropriate condition number for this problem.

<u>Conformal mapping</u> -- a mapping of one region upon another that is one-to-one and continuous, and such that angles ~~and distances~~ are preserved. Thus let arcs C_1 and C_2 in one region intersect in a point P; and let these be mapped into arcs C_1' and C_2' intersecting in P'. Let P_1 be any point on C_1, mapped into P_1' on C_1'. ~~Preservation of distances means that the arc length from P' to P_1' is the same as that from P to P_1.~~ Let φ be the angle between tangents to C_1 and C_2 at P, φ' that between tangents to C_1' and C_2' at P'. Preservation of angles means that $\varphi' = \varphi$. In case the regions are two-dimensional and simply connected, they can be regarded as regions of two complex planes, and the mapping is then defined by $z' = f(z)$, where $f(z)$ is analytic in z throughout the region.

Conjugate gradients, method of -- a method of successive approximation for solving a system of linear equations which ~~converges~~ is exact after at most n steps. The successive approximations are so chosen that each residual is orthogonal to all preceding ones.

Continuant -- a determinant of the form

$$K\begin{pmatrix} & a_1 \cdots a_n \\ b_0 & b_1 \cdots b_n \end{pmatrix} = \begin{vmatrix} b_0 & a_1 & 0 & \cdots & 0 \\ -1 & b_1 & a_2 & \cdots & 0 \\ 0 & -1 & b_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & b_n \end{vmatrix},$$

~~socalled~~ so called because the nth numerator and nth denominator of the continued fraction

$$F = b_0 + a_1/\overline{b_1} + \cdots$$

are

$$A_n = K\begin{pmatrix} & a_0 \cdots a_n \\ b_0 & b_1 \cdots b_n \end{pmatrix}$$

$$B_n = K\begin{pmatrix} & a_1 \cdots a_n \\ b_1 & b_2 \cdots b_n \end{pmatrix}.$$

Continued fractions -- Consider the (possibly infinite) system of difference equations

$$x_0 = b_0\, x_1 + a_1\, x_2,$$
$$x_1 = b_1\, x_2 + a_2\, x_3,$$
$$x_2 = b_2\, x_3 + a_3\, x_4,$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

So long as no x_i vanishes or becomes infinite one can write

$$x_0 / x_1 = b_0 + a_1 / (x_1 / x_2),$$
$$x_1 / x_2 = b_1 + a_2 / (x_2 / x_3),$$
$$x_2 / x_3 = b_2 + a_3 / (x_3 / x_4),$$

and with repeated substitution

$$x_0 / x_1 = b_0 + a_1 \overline{/ b_1} + a_2 \overline{/ b_2} + \ldots$$

where each horizontal bar signifies the inclusion of all that follows. The right-hand number is a continued fraction. If one arbitrarily sets $x_{i+1} = 0$ and $x_i = 1$, considering only the first i equations, the finite continued fraction can be evaluated by solving the ith equation for x_{i-1}; substituting in the preceding and solving for x_{i-2} / x_{i-1}; and continuing until x_0 / x_1 is obtained. In many applications the a_i and b_i are functions of a variable x, and one truncates and evaluates in the manner just described.

A finite continued fraction

$$F_n = b_0 + a_1 \overline{/ b_1} + \ldots + a_n / b_n \equiv A_n / B_n,$$

where A_n and B_n are polynomials in the a_i and b_i can be otherwise evaluated by means of the recursions (see <u>continuant</u>)

$$A_n = b_n A_{n-1} + a_n A_{n-2},$$
$$B_n = b_n B_{n-1} + a_n B_{n-2}.$$

These relations, however, though of great theoretical importance, are not to be recommended for computation in general since they tend to be highly unstable numerically.

The fraction F_n is called the nth approximant to F; A_n and B_n the nth numerator and denominator. If all A_n and B_n are given, and if

$$A_n / B_n \neq A_{n+1} / B_{n+1}$$

for every n, then the a_n and b_n are determined uniquely. In particular the

approximants F_{2n} are called the even parts, F_{2n+1} the odd parts, of F, and it is possible to form a continued fraction whose approximants are the even parts of F, and one whose approximants are the odd parts.

For applications of continued fractions, see reciprocal differences, Padé table.

Correction -- A quantity added to a previously obtained approximation to yield a better approximation. Sometimes the term is applied to $x - x^*$, where x is the true value and x^* the approximation, and is contrasted with the difference $x^* - x$ which is then designated the error. This sense is not the one adopted here.

Corrector -- In solving initial value problems in ordinary differential equations, a quadrature formula to be used for improving a given approximate value of the dependent variable at an advance point, and contrasted with a predictor used to obtain a first approximation. A predictor formula uses an open quadrature formula, a corrector formula a closed one.

Cracovians -- except for the rule of composition, the same as matrices, the Cracovian product of A and B being the matrix product $A^T B$. Advocated by Banachiewicz, they are in favor mostly among some European geodesists. Since any Cracovian identity can be expressed with matrices and conversely, it is clear that preference for either over the other can be based only on convenience, tradition, and taste, in some combination.

Cramer's rule -- in the system of linear equations

$$A x = h,$$

let $\delta = \det A$, and let δ_i represent the determinant obtained when h replaces the ith column of A. Then δ_i / δ are the elements of the solution vector x. The rule is of great theoretical importance but of little value in computing practice.

Crout's method -- a method of triangularization for matrix inversion.

Cubature formulas -- see multivariate interpolation.

Curve fitting -- Suppose that a physical variable y is known to bear a particular functional relation to another variable x:

$$y = f(x; \alpha, \beta, \ldots),$$

where the form of f is known, but where f contains certain parameters α, β, ..., that are unknown. Thus in radioactive decay the amount of a constituent is known to be an exponential function of time, $y = \alpha \exp(-\beta t)$, where t takes the place of the above x. If it is possible to measure the y associated with any x, one can make a series of measurements, x_i, y_i, and hope to solve the equations

$$y_i = f(x_i; \alpha, \beta, \ldots)$$

for the parameters. Since, in general, the measurements will be subject to error one must make more measurements than there are parameters, and since the equations will then be inconsistent, in general, it is necessary to establish some criterion for making the "best" selection of the parameters so that, in some sense, all equations are satisfied as nearly as possible, even though possibly none is satisfied exactly.

A possible criterion is the minimax criterion: to select the parameters so as to minimize the greatest of the quantities

$$|y_i - f(x_i; \alpha, \beta, \ldots)|.$$

This is sometimes used, but is very difficult computationally in most cases. Sometimes one resorts to trial and error, with visual inspection of the graphs. The method of least squares can often be justified on grounds of probability, and usually leads to the simplest computations.

The same mathematical problem arises when a known function $\varphi(x)$ is to be approximated in some fashion by a simpler one. For this purpose interpolation is

the most usual method of procedure.

Sometimes one has only the empirical measurements x_i, y_i, but no theory to prescribe a particular functional form $f(x; \ldots)$. In that case one can only apply intuition, and possibly trial and error, in selecting a convenient form. See smoothing. Ni.

Dandelin's method -- a root-squaring method for solving algebraic equations in which the equation $\varphi(x) = 0$ is replaced by

$$\varphi(\sqrt{x})\ \varphi(-\sqrt{x}) = 0,$$

differing only mildly from the later Graeffe's method. Dandelin applied this first in conjunction with Newton's method for accelerating convergence, but also proposed repeated application as a method in itself.

Decimal -- characterized by ten; compounded of ten parts. The decimal system of number representation is the ordinary one using the radix 10.

Deferred approach to the limit -- let φ represent a quantity to be calculated by a finite difference method (e.g., an integral); let φ_h represent the approximation obtained when $\Delta x = h$. Often it is approximately true that

$$\cancel{\varphi_h = \varphi + \varphi_1 h^2 + \varphi_2 h^4 + \ldots} \quad . \quad \phi_h = a_0 + a_1 h^2 + a_2 h^4 + \ldots ".$$

Let φ_{h_1} and φ_{h_2} represent similarly values obtained with $h = h_1$ and $h = h_2$. If h_1 and h_2 are small enough so that terms beyond h^2 can be neglected, then

$$\phi_{h_1 h_2} \equiv \left(h_1^2\ \varphi_{h_2} - h_2^2\ \varphi_{h_1}\right) \Big/ \left(h_1^2 - h_2^2\right)$$

is a much better approximation to φ than is either φ_{h_1} or φ_{h_2}. Commonly one takes $h_2 = h_1 / 2$.

Deflation -- a transformation applied to a matrix A when one proper value and vector are known for the purpose of reducing the computations required for finding the others. The general theorem is, if proper values of A are $\lambda_1, \lambda_2, \ldots, \lambda_n$, and if

$A\ v_1 = \lambda_1\ v_1$, then proper values of

$$A_1 = A - v_1\ k^T$$

are $\lambda_1 - k^T\ v_1, \lambda_2, \ldots, \lambda_n$, whatever the vector k. By choosing k^T to be a multiple of the first row of A, and scaled so that

$$k^T\ v_1 \equiv \lambda_1,$$

the first row of A_1 is null.

Delta-square process -- Let a_i, a_{i+1}, a_{i+2} represent three consecutive terms in a sequence, possibly three consecutive approximations to a root of an equation obtained by Newton's method, Bernoulli's method, or some other. In many situations, when the sequence has a limit, then $a_i - (\Delta\ a_i)^2\ /\ \Delta^2\ a_i$ will be much closer to this limit than will a_{i+3}. Here Δ represents the difference operator. If the approach is geometric, that is, if $a_i = a + \alpha\ \rho^i$, then this gives the limit a exactly.

The formula can be derived by assuming the above form for a_i and solving for a. A somewhat more general formula in terms of a_i, a_{i+1}, a_{i+2}, a_{i+3}, a_{i+4} can be obtained on the assumption that $a_i = a + \alpha_1\ \rho_1^i + \alpha_2\ \rho_2^i$. The result has a meaning even when $|\rho| \geq 1$, hence the method is sometimes effective for summing divergent series. Generalizations have been developed by Shanks and by Wynn.

Descartes's rule of signs -- for polynomials, a special case of Budan's theorem.

Detection of tabular errors -- by means of differences, is based upon the following table, readily verified and extended:

y	δy	$\delta^2 y$	$\delta^3 y$
0		0	
	0		ϵ
0		ϵ	
	ϵ		-3ϵ
ϵ		-2ϵ	
	$-\epsilon$		3ϵ
0		ϵ	
	0		$-\epsilon$
0		0	

This shows that the effect of small errors becomes accentuated progressively in the columns of differences.

Diagonal matrix -- see matrix.

Difference equation -- any equation involving finite differences. The usual methods for solving differential equations reduce to the solution of difference equations, with either the derivatives or the integrals expressed approximately in terms of finite differences (see differentiation formulas, quadrature formulas, ordinary differential equations, partial differential equations).

Difference operators -- For interpolation based upon equally spaced points, most standard formulas, as well as others, can be developed and expressed easily in terms of a few difference operators which can be manipulated formally as algebraic quantities. In their definition given below it is to be understood that they have no existence independent of the functions upon which they operate. Understanding that $\Delta x = h$ is fixed, the operators are defined as follows:

$E^u f(x) = f(x + u h)$ [displacement operator];

$\Delta f(x) = f(x + h) - f(x)$ [forward difference operator];

$\nabla f(x) = f(x) - f(x - h)$ [backward difference operator];

$\delta f(x) = f(x + h/2) - f(x - h/2)$ [central difference operator];

$2\mu f(x) = f(x + h/2) + f(x - h/2)$ [central mean operator];

$D f(x) = d f(x) / dx$ [differential operator];

$\theta f(x) = h d f(x) / dx.$

The last operator is not in common use, but simplifies the writing of some formulas and may be called the logarithmic displacement operator. For functions of several variables, subscripts can be used to specify the variable with respect to which the operator acts. Thus

$$E_x f(x, y, \ldots) = f(x + h, y, \ldots).$$

All operators can be defined formally in terms of one of them. Thus

$$\Delta = E - 1, \qquad \nabla = 1 - E^{-1}, \qquad E = \exp \theta,$$

the last being the operational expression of the Taylor series.

Examples of the application of these relations are to be found under appropriate headings, but the following illustration may be in order here. Let

$$x = x_0 + u\,h, \qquad u = (x - x_0) / h, \qquad f(x) = E^u f(x_0).$$

Since $E = 1 + \Delta$, therefore, by the binomial theorem,

$$E^u = 1 + \binom{u}{1} \Delta + \binom{u}{2} \Delta^2 + \binom{u}{3} \Delta^3 + \ldots$$

and therefore, formally, taking $x_i = x_0 + i\,h$, one has

$$f(x) = f(x_0) + \frac{x - x_0}{h} \Delta f(x_0) + \frac{(x - x_0)(x - x_1)}{2!\,h^2} \Delta^2 f(x_0) + \ldots$$

Unless u is a non-negative integer this is an infinite series, but for purposes of interpolation the series is truncated by dropping all but the first several terms. If all terms are dropped beyond that containing $\Delta^n f(x_0)$, then there is retained on the right that nth degree polynomial that passes through the n + 1 points whose coordinates are

$$x_i,\ f(x_i), \qquad i = 0, 1, \ldots, n,$$

and the formula is <u>Newton's interpolation formula</u> with forward differences. If one writes $E = (1 - \nabla)^{-1}$ and proceeds similarly one obtains <u>Newton's interpolation formula</u> with backward differences:

$$f(x) = f(x_0) + \frac{x - x_0}{h} \nabla f(x_0) + \frac{(x - x_0)(x - x_{-1})}{2!\,h^2} \nabla^2 f(x_0) + \ldots$$

It is to be noted that the binomial coefficients

$$\binom{u}{r} = \frac{u(u-1) \ldots (u-r+1)}{r!}$$

are defined even when u is not a positive integer, and they are polynomials in u. Sometimes the symbol $u_{(r)}$ is used, and related to these polynomials are the factorial polynomials

$$u^{(r)} = u(u - 1) \ldots (u - r + 1) = r!\, u_{(r)}.$$

Note that

$$\Delta u^{(r)} = r\, u^{(r-1)}, \qquad \Delta u_{(r)} = u_{(r-1)}.$$

Differential analyzer -- a computer, generally analog, designed for the purpose of solving differential equations.

Differential equation -- an equation which relates, for each choice of the independent variable (s), the values of one or more dependent variable and one or more derivative of each. It is an ordinary differential equation in case only a single independent variable occurs; a partial differential equation when there are two or more. For a dependent variable to be uniquely defined as a function of the independent variable (s), there must be associated a sufficient number of initial or boundary conditions, or a combination of both. For methods of numerical solution see ordinary differential equations, partial differential equations, elliptic equations, parabolic equations, hyperbolic equations.

Differentiation formulas (numerical) -- a quite general procedure for carrying out numerical differentiation is to use interpolation, or least squares, or some other method for approximating the function of interest, and then to differentiate the approximating function. When the points of interpolation are unequally spaced one can use divided differences to form an interpolation polynomial.

When the points are uniformly spaced, a number of convenient formulas can be derived by means of <u>difference operators</u>. Since

$$h\,D = \theta = \log\,(1 + \Delta) = -\log\,(1 - \nabla),$$

one obtains

$$\begin{aligned}\theta &= \Delta - \Delta^2 / 2 + \Delta^3 / 3 - \ldots \\ &= \nabla + \nabla^2 / 2 + \nabla^3 / 3 + \ldots,\end{aligned}$$

in terms of forward and backward differences, respectively. Derivatives of higher order can be obtained by raising to higher powers the expressions for θ.

Expressions in terms of central differences require more manipulation to obtain, but one finds

$$\theta = \mu[\delta - \delta^3/3! + (2!)^2\,\delta^5/5! - (3!)^2\,\delta^7/7! + \ldots],$$
$$\theta^2 = 2[\delta^2/2! - \delta^4/4! + (2!)^2\,\delta^6/6! - (3!)^2\,\delta^8/8! + \ldots].$$

To obtain higher powers write $\theta^3 = \theta \cdot \theta^2$, $\theta^4 = (\theta^2)^2$, $\theta^5 = \theta \cdot \theta^4$, ... To apply, one truncates the series at some point, and if desired one can replace Δ, ∇, δ and $\mu\,\delta$ by their expressions in terms of E.

In case the derivative is required at a point other than a tabulated one, say at $x_0 + u\,h$, one must apply E^u, expressing E in terms of the appropriate difference operator. However, the error can be expected to be substantially greater. See <u>Sterling numbers</u>.

<u>Digital computer</u> -- As contrasted with <u>analog computer</u>, a computing aid that represents numbers digitally, in decimal, binary, or (conceivably) other base. The simplest artificial example is the abacus, which represents numbers digitally in such a way as to facilitate carrying out additive operations, hence multiplication and division as sequences of these. Desk computers perform these operations automatically, and sometimes also square-rooting or other special sequences. The very high speed, general purpose computers are <u>stored-program computers</u> and are digital.

Direct methods -- usually contrasted with iterative methods or methods of successive approximation for matrix inversion and the solution of equations. The exact inverse or solution is expressed as the result of a finite sequence of operations or scalars.

Displacement operator -- see difference operators.

Divided differences -- For interpolation with unequally spaced abscissae, let $f(x)$ be the function to be interpolated, and let its values be given at

$$x_0, x_1, \ldots, x_n,$$

where it is not required that they be in order or even distinct. Form the table as follows:

x_0	$f(x_0)$				
		$f(x_0, x_1)$			
x_1	$f(x_1)$		$f(x_0, x_1, x_2)$		
		$f(x_1, x_2)$		$f(x_0, x_k, x_2, x_3)$	
x_2	$f(x_2)$		$f(x_1, x_2, x_3)$		....
		$f(x_2, x_3)$		$f(x_1, x_2, x_3, x_4)$	
x_3	$f(x_3)$		$f(x_2, x_3, x_4)$		
		$f(x_3, x_4)$			
x_4	$f(x_4)$				

where

$$f(x_i, x_j) = [f(x_i) - f(x_j)] / (x_i - x_j);$$
$$f(x_i, x_j, x_k) = [f(x_i, x_j) - f(x_j, x_k)] / (x_i - x_k);$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

When there are coincident abscissae they should be placed in adjacency, and thus

$$f(x, x) = f'(x),$$
$$f(x, x, x) = f''(x) / 2!,$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

It can be shown that any divided difference is a symmetric function of all its arguments, which means that

$$f(\ldots, x_i, \ldots, x_j, \ldots) = f(\ldots, x_j, \ldots, x_i, \ldots),$$

it being understood that dots represent arguments that are the same on the two sides.

In terms of divided differences one can write the following interpolation formula due to Newton:

$$f(x) = f(x_0) + (x - x_0)\, f(x_0, x_1) + (x - x_0)\,(x - x_1)\, f(x_0, x_1, x_2)$$
$$+ \ldots + (x - x_0) \ldots (x - x_n)\, f(x_0, x_1, \ldots, x_n, x),$$

where the last term written is the <u>remainder</u> and previous ones constitute the interpolation formula.

<u>Doolittle method</u> -- a particular scheme of recording intermediate results in applying triangularization to solve a system of normal equations. See <u>matrix inversion</u>.

<u>Dot product</u> -- see <u>scalar product</u>.

<u>Eigenfunction</u> (Eigenfunktion) -- see <u>eigenvalue</u>.

<u>Eigenvalue</u> (Eigenwert) -- any value of the scalar λ for which the integral equation $\varphi(x) = \lambda \int K(x, y)\, \varphi(y)\, d\, y$ has a nontrivial solution. The associated solution $\varphi(x)$ is called the eigenfunction (from the German Eigenfunktion). The term is also commonly applied to proper value of a <u>matrix</u>, but if it is to be used in this connection it should denote the reciprocal of the proper value. But usage varies even for integral equations.

<u>Eigenvector</u> -- a proper vector of a matrix. See <u>eigenvalue</u>, <u>matrix</u>, <u>principal vector</u>.

<u>Elimination</u> -- a method of solving a system of equations in which one of the equations is solved for one of the unknowns in terms of the others, and the solution used to replace this unknown in all the other equations. Thus if the orig-

inal system was one of n equations in n unknowns, one obtains a new system of which n - 1 of the equations contain only n - 1 of the unknowns.

Elliptic equation -- a partial differential equation of particular form. Ordinarily, along with the equation, it is necessary if the solution is to be well defined to impose conditions along the boundary, such as the value of the dependent variable (Dirichlet condition), or its normal derivative (Neumann condition). In the simple case the ~~Poisson~~ Laplace operator $D_x^2 + D_y^2$ is represented by $h^{-2}(\delta_x^2 + \delta_y^2)$ which gives rise to the coefficient scheme

$$\begin{array}{ccc} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{array}$$

which is always implicit (see matrix inversion). Refer to Co, Ku, for more details on methods.

Enlargement, methods of -- methods of matrix inversion by which the inverse of a matrix of order m is expressible in terms of that of a submatrix of order m - 1.

Error -- in general, any deviation of a computed quantity from the theoretically correct value. In a restricted sense, that deviation due to unavoidable random disturbances, or to the use of finite approximations to what is defined by an infinite series, and hence to be contrasted with blunder. Errors of measurement arise from the use of instruments of less than infinite precision, and from random, uncontrollable disturbances. Truncation errors are due to the use of a term as an approximation for the limit of an infinite sequence, or, in general, to the use of any approximation in place of the theoretically correct quantity (cf Remainder, Roundoff). If f(x) is to be computed but only the approximation x* to x is known, then f(x) - f(x*) is the propagated error. Bounds for the propagated error can be had by applying mean value theorems when f has suitable

continuity properties. Generated errors are those which arise in the computation itself because of rounding. They depend upon the details of the routine.

If f is the true and f* the computed value, sometimes f - f* is called the correction, f* - f the error. The term correction, however, will be used in a different way here, to designate a quantity to be added to f* for reducing the error, hence an approximation to the error.

Escalator method -- a method of enlargement for matrix inversion, and for computing proper values and vectors of matrices. Euclidean algorithm -- see higest common factor.

Euler constant --

$$\gamma = \lim_{n \to \infty} \left(\sum_{1}^{n} k^{-1} - \log n \right)$$

$$= \cancel{0.577215665\ldots}. \quad 0.577215665.$$

Euler-Maclaurin formula -- the formal expansion.

$$\int_{x_0}^{x_m} f(x)\, dx = h(f_0/2 + f_1 + f_2 + \ldots + f_{m-1} + f_m/2)$$
$$- \sum_{2}^{\infty} B_\nu\, h^\nu (f_m^{(\nu-1)} - f_0^{(\nu-1)}) \,/\, \nu!,$$

where

$$f_i = f(x_i) = f(x_0 + i\, h)$$

and the B_ν are the Bernoulli numbers.

Euler method -- for solving an ordinary differential equation

$$y' = f(x, y)$$

takes

$$y_{\nu+1} = y_\nu + h\, f_\nu,$$
$$f_\nu = f(x_\nu, y_\nu).$$

The "improved Euler method" obtains sequentially

$$y_{\nu+\frac{1}{2}} = y_\nu + h\, f_\nu / 2,$$
$$y_{\nu+1} = y_\nu + h\, f_{\nu+\frac{1}{2}}$$

and can be considered an elementary form of the Runge-Kutta method. The "improved Euler-Cauchy method" forms

$$y^*_{\nu+1} = y_\nu + h\, f_\nu, \quad f^*_{\nu+1} = f(x_{\nu+1}, y^*_{\nu+1}),$$
$$y_{\nu+1} = y_\nu + h(f_\nu + f^*_{\nu+1}) / 2.$$

See Co.

Euler summation formula -- the Euler-Maclaurin formula when used for summing a series.

Euler transformation -- of an infinite series whose partial sums are

$$S_n = a_0 + a_1 x + \ldots + a_n x^n$$

utilizes the identity

$$(1 - x) S_n = a_0 - a_n x^{n+1} + x \sum_0^{n-1} \Delta a_i x^i.$$

It is a method for the summation of series.

Everett interpolation formula -- in the notation of difference operators

$$E^u = (1-u) + \binom{2-u}{3} \delta^2 + \binom{3-u}{5} \delta^4 + \ldots + \left[u + \binom{1+u}{3} \delta^2 + \binom{2+u}{5} \delta^4 + \ldots\right] E.$$

It thus requires differences of even orders only, but they both must be taken at x_0 and at $x_1 = x_0 + h$.

Extrapolation -- contrasted with interpolation when the function is required at a point outside the range of the x_i. The same methods are applied.

Factor theorem -- see algebraic equations.

Factorial coefficients -- Stirling numbers, so called from their definition.

False position, method of -- see regula falsi.

Fixed-point -- a system of computing in which all quantities must be scaled so as to be representable by a fixed number of digits and with the radix point in a fixed position. Contrasted with floating-point computations in which a scale factor is explicitly computed and associated with each number, and represented as a power of the radix. In some digital computers floating-point operations are automatic; in others they are not.

Floating-point -- see fixed-point.

Forward-difference operator -- see difference operators.

Forward differences -- see difference operators.

Forward interpolation -- see Newton interpolation formulas.

Fourier series -- a general series of the form

$$f(x) = a_0 + a_1 \cos x + a_2 \cos 2x + \ldots + b_1 \sin x + b_2 \sin 2x + \ldots$$

is called a trigonometric series.

Because of the orthogonality relations

$$\int_{-\pi}^{\pi} \cos px \cos qx \, dx = \int_{-\pi}^{\pi} \sin px \sin qx \, dx = 0, \quad p \neq q,$$

$$\int_{-\pi}^{\pi} \cos px \sin qx \, dx = 0$$

together with

$$\int_{-\pi}^{\pi} \cos^2 px \, dx = \int_{-\pi}^{\pi} \sin^2 px \, dx = \pi,$$

it follows that

$$a_p = \pi^{-1} \int_{-\pi}^{\pi} f(x) \cos px \, dx, \quad b_p = \pi^{-1} \int_{-\pi}^{\pi} f(x) \sin px \, dx,$$

and with a_p and b_p so defined the series is called a Fourier series.

Fredholm equation -- an integral equation with fixed limits:

$$f(x) = \varphi(x) - \lambda \int_a^b K(x, \xi) \varphi(\xi) \, d\xi$$

where φ is unknown. For a numerical solution, let $a \leq \xi_0 \leq \ldots \leq \xi_n \leq b$ and apply a quadrature formula of the form

$$\int_a^b g(x) \, dx \doteq \sum_0^n \eta_i \, g(\xi_i).$$

Then

$$f(\xi_i) = \varphi(\xi_i) - \lambda \sum_0^n \eta_j \, K(\xi_i, \xi_j) \, \varphi(\xi_j).$$

These are n + 1 equations in n + 1 unknowns $\varphi(\xi_j)$ (matrix inversion). See Co, Ko.

Functional iteration -- see iterative methods.

Galerkin method -- see Ritz method.

Gaussian elimination -- matrix inversion by triangularization.

Gaussian interpolation formulas -- (also called the Newton-Gauss formulas) in the notation of difference operators, the forward formula is

$$E^u = 1 + u\,\delta\,E^{\frac{1}{2}} + \binom{u}{2}\delta^2 + \binom{u+1}{3}\delta^3\,E^{\frac{1}{2}} + \binom{u+1}{4}\delta^4 + \ldots$$

and the backward formula is

$$E^{-u} = 1 - u\,\delta\,E^{-\frac{1}{2}} + \binom{u}{2}\delta^2 - \binom{u+1}{3}\delta^3\,E^{-\frac{1}{2}} + \binom{u+1}{4}\delta^4 - \ldots$$

Gaussian quadrature formula -- any formula of the form

$$\int_a^b w(x)\,f(x)\,dx = \sum_0^n \mu_i\,f(x_i) + R$$

where the density function $w(x) \geq 0$; the coefficients μ_i are constant and depend only upon the function $w(x)$ and the range of integration; and the abscissae x_i are selected so that the remainder R vanishes whenever $f(x)$ is a polynomial of degree 2n + 1 or less. It can be shown that these x_i exist and satisfy an algebraic equation of degree n + 1; that they are real; and that when ordered they satisfy $a < x_0 < x_1 < \ldots < x_n < b$. They are, in fact, zeros of one of a set of orthogonal polynomials.

If a and b are both finite, one can make a substitution of variable, if necessary, and suppose $b = -a = 1$. Then if $w(x) = 1$, the x_i are zeros of the Legendre polynomial of degree n + 1, and one has for

$n = 1$: $\mu_0 = \mu_1 = 1$;

$n = 2$: $\mu_0 = \mu_2 = 5/9$, $\mu_1 = 8/9$;

$n = 3$: $\mu_0 = \mu_3 = (18 - \sqrt{30}) / 36$, $\mu_1 = \mu_2 = (18 + \sqrt{30}) / 36$.

By choosing $w(x)$ appropriately either a or b or both can be infinite. For tabulation of x_i and μ_i, and for a discussion of the theory, see Ko.

Gauss-Seidel method -- a term sometimes applied to the total step iteration for solving linear equations (see matrix inversion) although both Gauss and Seidel used a relaxation method.

Generated error -- see error.

Givens method -- for computing proper values of a hermitian matrix.

Graduation -- another term for smoothing.

Graeffe method -- a root-squaring method for solving algebraic equations. Given the algebraic equation, separate odd and even powers of x on opposite sides of the equation:

$$x^n + a_2 x^{n-2} + a_4 x^{n-4} + \ldots = a_1 x^{n-1} + a_3 x^{n-3} + \ldots.$$

If both sides are squared, all powers will be even, and x^2 can be replaced by y. Separate in the same way and repeat. After p such operations, let the equation be

$$z^n + c_2 z^{n-2} + \ldots = c_1 z^{n-1} + \ldots.$$

Then the roots z_i of this equation, and the roots x_i of the first, are related by

$$z_i = x_i^{\nu}, \qquad \nu = 2^p.$$

Then if

$$|x_1| > |x_2| \geq \ldots,$$

it follows that, approximately,

$$c_1 \doteq x_1^{\nu}.$$

If

$$|x_1| \geq |x_2| > |x_3| \geq \ldots,$$

then, approximately,

$$c_2 \doteq x_1^{\nu} x_2^{\nu}.$$

And, if

$$|x_1| = |x_2| > |x_3| \geq \ldots,$$

then $z_1 = x_1^{\nu}$ and $z_2 = x_2^{\nu}$ satisfy, approximately,

$$z^2 - c_1 z + c_2 \doteq 0.$$

When all roots are distinct,

$$x_2^{\nu} \doteq c_2/c_1, \quad x_3^{\nu} \doteq c_3/c_2, \quad x_4^{\nu} \doteq c_4/c_3, \ \ldots$$

approximately. An enumeration of all possible contingencies resulting from the presence of conjugate pairs equal in modulus is out of the question, but generally speaking, a term that continues to oscillate rather than to increase after continued squaring marks the presence of a conjugate pair, and the pair z_i and z_{i+1} will satisfy approximately

$$c_{i-1} z^2 - c_i z + c_{i+1} = 0$$

(see <u>symmetric functions</u>). The method can be generalized to apply to a function expressible in the form

$$f(x) = 1 + a_1 / x + a_2 / x^2 + \ldots.$$

If the series converges for $|x| > R$, then all zeros of $f(x)$ exceeding R in modulus can be found (see Ho).

<u>Gram-Schmidt orthogonalization</u> -- see <u>orthogonalization</u>.

Green's function -- a function $K(x, \xi)$ associated with a given boundary value problem $L[y] = \varphi(x)$ that has the property that the solution is expressible in the form

$$y(x) = \int K(x, \xi)\, \varphi(\xi)\, d\xi,$$

the integration extending over the interval on which the solution is to be defined.

As an example, if $L[y] \equiv y''$, and the boundary conditions are $y(0) = y(1) = 0$, then

$$K(x, \xi) = \begin{cases} (1 - \xi)\, x & \text{for } x \leq \xi, \\ (1 - x)\, \xi & \text{for } x > \xi. \end{cases}$$

The limits of integration are then from 0 to 1. The notion can be generalized to spaces of higher dimensionality. For theoretical investigations the function is of fundamental importance, but for numerical computation it is seldom used.

Gregory formula -- for numerical quadrature:

$$\int_{x_0}^{x_r} f(x)\, dx = h(\tfrac{1}{2} f_0 + f_1 + f_r + \dots + \tfrac{1}{2} f_r)$$
$$- (\Delta f_{r-1} - \Delta f_0) / 12 - (\Delta^2 f_{r-2} + \Delta^2 f_0) / 24$$
$$- 19(\Delta^3 f_{r-3} - \Delta^3 f_0) / 720 - 3(\Delta^4 f_{r-4} + \Delta^4 f_0) / 160 + \dots$$

Gregory-Newton formula -- generally called the Newton formula for forward differences, though known to Gregory (see difference operators).

Group relaxation -- a variant of the total step iteration for solving a linear system (see matrix inversion) in which at each step two or more equations are satisfied simultaneously. Same as block relaxation.

Hermite interpolation formula -- makes use of the polynomial of degree $2n + 1$ with prescribed value and prescribed slope at each of the $n + 1$ abscissae

$$x_0, x_1, \dots, x_n,$$

assumed distinct. The polynomial is

$$H(x) = \sum_{0}^{n} [f_i \; v_i(x) + f_i' \; w_i(x)] \; L_i^2(x),$$

where f_i and f_i' are the prescribed values and slopes; where

$$w_i(x) = x - x_i, \quad v_i(x) = 1 - w_i(x) \; \omega''(x_i) \; / \; \omega'(x);$$

and where the $L_i(x)$ and $\omega(x)$ occur in the Lagrange interpolation formula.
More generally, higher derivatives may also be prescribed.

Hermitian matrix -- a matrix A such that $A^* = A$. Its proper values are all real and it can always be diagonalized. In fact, there exists a unitary matrix V such that

$$A = V \Lambda V^*$$

where Λ is diagonal. The computation of the proper values and vectors is much simpler, therefore, than in the more general case.

One of the most successful methods is that of Givens, which will be described for the symmetric (real) case: Let R_{23} be a matrix that differs from the identity only in the elements (2, 2), (3, 3), (2, 3), and (3, 2), and let

$$\rho_{22} = \cos \varphi = \rho_{33}; \quad \rho_{23} = -\rho_{32} = \sin \varphi.$$

Then R_{23} is orthogonal, and φ can be chosen so that in

$$R_{23}^T \; A \; R_{23}$$

the elements in positions (1, 3) and (3, 1) vanish. If a matrix R_{24} is formed analogously, the elements in (1, 4) and (4, 1) can be annihilated without disturbing the zeros previously created. By continuing this process, the final matrix will be tridiagonal in form:

$$T = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & 0 & \cdots \\ \beta_1 & \alpha_2 & \beta_2 & 0 & \cdots \\ 0 & \beta_2 & \alpha_3 & \beta_3 & \cdots \end{pmatrix}.$$

The characteristic polynomial can be obtained by means of a very simple recursion:

$$\varphi_0(\lambda) = 1,$$

$$\varphi_1(\lambda) = \lambda - \alpha_1,$$

$$\varphi_2(\lambda) = (\lambda - \alpha_2)\,\varphi_1(\lambda) - \beta_1^2\,\varphi_0(\lambda),$$

$$\varphi_3(\lambda) = (\lambda - \alpha_3)\,\varphi_2(\lambda) - \beta_2^2\,\varphi_1(\lambda),$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$\varphi_n(\lambda)$ being the required polynomial. However, if every $\beta_i \neq 0$ the polynomials $\varphi_i(\lambda)$ form a Sturm sequence, possessing all the properties required for satisfying the Sturm theorem. Consequently, for any μ, by counting the variations in sign of the sequence of values $\varphi_i(\mu)$ one can ascertain the precise number of proper values exceeding and exceeded by μ. It is therefore advantageous not to expand the polynomial $\varphi(\lambda)$ explicitly, but to proceed as follows for solving the equation: Having found an interval on which one or more roots are known to lie, take consecutive midpoints, evaluating the terms of the sequence for each by means of the recursion.

If the vectors are required, it will be necessary to form the product R of all the rotations applied. Then

$$A\,R = R\,T,$$

and if t is a proper vector of T belonging to λ:

$$T\,t = \lambda\,t,$$

then

$$A\,R\,t = R\,T\,t = \lambda\,R\,t$$

so that R t is a proper vector of A belonging to λ. (See also Lanczos method of biorthogonalization).

Another method, often called the Jacobi method, for computing proper values is iterative in character, and makes use of the fact that the matrix A can be trans-

formed to the diagonal form Λ as the limit of an infinite sequence of plane rotations. The rotation matrices R_{ij} are of the same form as before, but the angle φ is chosen so that in the matrix

$$R_{ij}^{T} \; A \; R_{ij}$$

the elements vanish in the positions (i, j) and (j, i). Zeros are not preserved, in general, but the sum of the squares of the diagonal elements is increased by the squares of the annihilated elements at each step. Optimally one should, at each step, annihilate the pair of largest off-diagonal elements, but to reduce the searching one can go through the elements above the diagonal in some sequential order, annihilating all elements greater than some specified magnitude. With successive cycles the level should be decreased. In the limit the product of all the rotation matrices R_{ij} is the matrix V of proper vectors. This method has been described in detail by von Neumann, Goldstine, and Murray.

<u>Hessenberg method</u> -- see <u>proper values and vectors</u>.

<u>Highest common factor</u> -- of two polynomials, P_0 and P_1, is the polynomial of greatest degree that divides both. In case this is a constant, the polynomials are said to be relatively prime. It can be found by applying the Euclidean algorithm: Supposing the degree of P_1 not exceeding that of P_0, divide P_1 into P_0 and let $-P_2$ represent the remainder; divide P_2 into P_1 with $-P_3$ the remainder; ...; then if $P_i \neq 0$, $P_{i+1} = 0$, P_i is the highest common factor. This is represented

$$P_0 = Q_1 \, P_1 - P_2,$$

$$P_1 = Q_2 \, P_2 - P_3,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$P_{i-1} = Q_i \, P_i.$$

Hitchcock method -- for finding complex roots of an algebraic equation. See Bairstow method.

Horner method -- for obtaining irrational roots of an algebraic equation: Given that a particular root is located between consecutive integers n_0 and $n_0 + 1$, reduce the roots by n_0. Then locate the root between consecutive tenths, say $n_1/10$ and $(n_1 + 1)/10$, $0 \leq n_1 \leq 9$, and reduce by $n_1/10$. By continuing, the root can be determined as

$$n_0 + n_1 \, 10^{-1} + n_2 \, 10^{-2} + \ldots$$

to as many decimals as desired (see synthetic division). For hand computation the method is not to be recommended, but with a binary machine the method takes the form of successive bisection of an interval on which a root is known to lie, in order to obtain successive binary digits, and the method is often very effective (see algebraic equations).

Hyperbolic equation -- a second order partial differential equation. For the solution to be well defined ordinarily initial and boundary conditions are needed. In the simple case the equation can be reduced to the form $u_{xx} - u_{yy} + \ldots = 0$, where omitted terms involve no derivatives of second or higher order. The usual difference scheme is representable

$$\begin{array}{ccc} & 1 & \\ \tau^2 & 2(1 - \tau^2) & \tau^2 \\ & 1 & \end{array}$$

where $\tau = \Delta y / \Delta x$. If initial conditions are given along a horizontal line, it is necessary that $\tau \leq 1$ for adequate numerical stability.

Initial-value problems -- for ordinary differential equations those for which the initial values of the dependent variables are specified; for a partial differential equation one of parabolic or hyperbolic type, requiring the specification of initial conditions as well as possibly conditions along the boundary.

Inner product -- a distributive operation defined on pairs of vectors or functions in some space to yield a scalar, and such that the result is real and non-negative when the two vectors or functions are the same. In the latter case it may be taken to represent the squared length, and thus a metric is defined.

For ordinary vectors the inner product is generally taken to be the ordinary scalar product. In function space one first prescribes a nondecreasing real function σ, and defines the inner product of φ and ψ to be

$$(\varphi, \psi) = \int \bar{\varphi}\, \psi \, d\, \sigma.$$

To say that the operation is distributive means that

$(\alpha_1 \varphi_1 + \alpha_2 \varphi_2, \psi) = \bar{\alpha}_1(\varphi_1, \psi) + \bar{\alpha}_2(\varphi_2, \psi)$, and $(\varphi, \beta_1 \psi_1 + \beta_2 \psi_2) = \beta_1(\varphi, \psi_1) + \beta_2(\varphi, \psi_2)$

where α_1, α_2, β_1, β_2 are scalars.

Integration, numerical -- see quadrature formulas.

Interpolation -- the ~~determination~~ approximation from tabulated values of a function, and possibly of its derivatives, of values not included in a table. Let $f(x)$, and possibly derivatives, be tabulated at $x_0, \ldots, x_m$. Usually one selects a set of functions $\varphi_0, \ldots, \varphi_n$, which may be polynomials, trigonometric functions, exponentials, or of any other convenient class, and applies some criteria for selecting coefficients α_i such that $\alpha_0 \varphi_0 + \ldots + \alpha_n \varphi_n$ is accepted as an approximation $P[f|x]$ to $f(x)$. The most common criteria are agreement of P with f (and possibly derivatives) at the x_i. Most commonly the term "interpolation" is applied only when criteria of this type are applied. Cf. difference operators; Aitken method; Lagrange interpolation formula; Hermite interpolation formula; Newton interpolation formula; and, for other criteria, minimax approximation, least squares.

Inverse interpolation -- the ~~determination~~ approximation, from an interpolating polynomial, of the value of an abscissa corresponding to a given ordinate. This amounts to solving an algebraic equation of degree equal to the polynomial. See Ko, W and R.

Iterative methods -- for solving equations, whether algebraic or transcendental, are, in fact, methods of successive approximation in which, having given one or more approximations to a solution, it is used in computing an improved one. Only the case of a single equation in a single variable will be considered here (see matrix inversion and the solution of linear equations, and simultaneous equations; also Ho for further development of the theory).

If the equation to be solved is

$$f(x) = 0,$$

let

$$\varphi(x) = x - g(x)\, f(x),$$

where throughout some region containing α, the root to be determined, $g(x)$ nowhere vanishes or becomes infinite. Then

$$\alpha = \varphi(\alpha),$$

and if, for some x_0 in this region, every

$$x_{i+1} = \varphi(x_i)$$

is again in the region and the sequence of x_i converges, it necessarily converges to a root. A sufficient condition for this is that

$$\cancel{|\varphi'(x)| < 1} \quad |\phi'(x)| \leq K < 1$$

at every point of the region. Moreover, if φ is analytic in some circle about x_0, and if it can be shown that for some positive $k < 1$,

$$|\varphi(x') - \varphi(x'')| < k|x' - x''|,$$

whenever both x' and x'' are in the circle, then it can be concluded that every x_i will, in fact, fall within the circle and that the equation has a root α to which

the sequence converges.

Newton's method is obtained with

$$g(x) = 1 / f'(x),$$

and one is assured of convergence if, for real α,

$$f(x_0)\ f''(x_0) > 0,$$

and neither f' nor f'' changes sign between α and x_0.

Jacobi method -- for computing proper values of a hermitian matrix.

Jordan method -- a method of matrix inversion in which by a succession of row operations one reduces the given matrix to diagonal form.

Jordan normal form -- see matrix.

Kronecker delta --

$$\delta_{ij} = \begin{matrix} 0 & \text{if} & i \neq j \\ 1 & \text{if} & i = j. \end{matrix}$$

Krylov method -- see proper values and vectors.

Lagrange interpolation formula -- a formula for interpolation that may be used when the abscissae $x_0, x_1, \ldots, x_n$ at which $f(x)$ is known are distinct but not necessarily equally spaced. Let

$$\omega(x) = (x - x_0)\ (x - x_1)\ \ldots\ (x - x_n),$$

$$L_i(x) = \omega(x) / [(x - x_i)\ \omega'(x_i)]$$
$$= \prod_{j \neq i}^{0,n} [(x - x_j) / (x_i - x_j)].$$

Then the Lagrange formula is

$$P[f|x] = \sum f_i\ L_i(x).$$

The polynomials $L_i(x)$ have been tabulated for particular spacings of the abscissae, since they are independent of the function f to be interpolated. In case some x_i are coincident, confluent forms exist (see Ho for a general theory), but in practice it is simpler then to use divided differences.

Laguerre polynomials -- orthogonal polynomials on the interval (0, ∞) with the density function

$$w(x) = x^{\alpha} e^{-x}, \quad \alpha > -1.$$

They satisfy the recursion

$$L_{n+1}(x) - (x - \alpha - 2n-1)\, L_n(x) + n(\alpha + n)\, L_{n-1}(x) = 0$$

with

$$L_0(x) = 1, \quad L_1(x) = x - \alpha - 1.$$

Lanczos method of biorthogonalization -- a method of forming sequentially the columns of matrices B and C such that $B^T C = D$ is diagonal, and, for a given matrix A,

$$A\,B = B\,T, \qquad A^T C = C\,T,$$

$$T = \begin{pmatrix} z_{11} & z_{12} & 0 & 0 & \cdots \\ 1 & z_{22} & z_{23} & 0 & \cdots \\ 0 & 1 & z_{33} & z_{34} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

a tridiagonal matrix. If the initial columns of B and C are chosen so that no diagonal element of D vanishes, the reduction can be carried through to completion barring special properties of A. See proper values and vectors. For Hermitian matrices the method is equivalent to that of Givens.

Laplace differential equation -- $\nabla^2 u = 0,$

where ∇^2 is the Laplacian operator. The nonhomogeneous form, in which a known function of the independent variables occurs on the right, is known as the Poisson equation.

Laplacian operator -- sometimes represented by Δ, sometimes by ∇^2, and of the form

$$\partial^2 / \partial x^2 + \partial^2 / \partial y^2 + \ldots.$$

See Laplace differential equation, Poisson differential equation.

Latent roots -- see proper values and vectors, matrix.

Least squares -- a method of smoothing or curve fitting which selects the fitted curve (or surface) so as to minimize the sum of the squares of the deviations of the given points from the curve (or surface). In the linear case one has a system which can be expressed in matrix form as

$$X\,c = y + d,$$

where the matrix X is N x n, with N > n; y is the vector of measured quantities; and the length of the residual vector d is to be minimized. If the measurements are equally weighted, this means that c is chosen so that $d^T d$ is as small as possible; if not, then $d^T W d$ is to be made as small as possible, where W is the diagonal matrix whose elements are the weights. The solution is that of the system

$$X^T W X x = X^T W y,$$

where W = I if the weights are equal (see matrix inversion). For fitting a polynomial the matrix X has the form

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \dots & x_3^{n-1} \\ 1 & x_4 & x_4^2 & \dots & x_4^{n-1} \\ . & . & . & . & . \end{pmatrix}$$

where x_i is the value of the independent variable for the ith measurement. However, the use of orthogonal polynomials is advisable for this problem.

In polynomial fitting the unknown parameters are the coefficients of the polynomial, and they occur linearly in the equations they satisfy. In case the theoretical relationship is of the form

$$\eta = f(\xi;\ \alpha,\ \beta,\ \dots)$$

where the parameters α, β, ..., do not occur linearly, it is generally necessary

to linearize the equations in order to solve them. Two methods can be suggested, and either requires initial, sufficiently close, approximations α_0, β_0, ..., which are to be improved progressively. By one method one forms the quantity to be minimized,

$$\delta(\alpha, \beta, \ldots) = \sum [\eta_i - f(\xi_i; \alpha, \beta, \ldots)]^2,$$

the summation extending over all observations,
and applies a method of steepest descent. In the other method one expands in Taylor series

$$f(\xi_i; \alpha_0, \beta_0, \ldots) + (\alpha - \alpha_0) f_\alpha(\xi_i; \alpha_0, \beta_0, \ldots) + (\beta - \beta_0) f_\beta(\xi_i; \alpha_0, \beta_0, \ldots) + \ldots = \eta_i + \delta_i.$$

Here one wishes to minimize $\sum \delta_i^2$, or possibly $\sum \omega_i \delta_i^2$ where the ω_i are weights. If α_0, β_0, ... are sufficiently close to the true α, β, ..., one can omit quadratic and higher terms in $\alpha - \alpha_0$, $\beta - \beta_0$, ..., and solve as a linear least squares problem for these differences. The result could be added as partial corrections to α_0, β_0, ..., and the process repeated. See De.

Legendre polynomials -- orthogonal polynomials on the interval (-1, 1) with the ~~weight~~ density function $w(x) = 1$. They satisfy the recursion

$$(n + 1) P_{n+1}(x) - (2n + 1) x P_n(x) + n P_{n-1}(x) = 0,$$

and

$$P_0 = 1, \quad P_1 = x, \quad P_2 = (3x^2 - 1) / 2,$$

$$P_3 = (5x^3 - 3x) / 2, \quad P_4 = (35x^4 - 30x^2 + 3) / 8.$$

Liebmann method -- the Gauss-Seidel method applied to the solution of the difference equations approximating the Laplace or the Poisson equation. See matrix inversion, partial differential equations.

Linear equations (algebraic) -- equations in which the unknowns occur only linearly, hence of the form

$$a_{11} x_1 + a_{12} x_2 + \dots = h_1,$$
$$a_{21} x_1 + a_{22} x_2 + \dots = h_2.$$

They are homogeneous if $0 = h_1 = h_2 = \dots$, otherwise nonhomogeneous. In matrix form the equations can be written

$$A\ x = h,$$

and the condition for solvability is that the rank of A and the rank of the augmented matrix (A, h) be the same. For methods of solving see matrix inversion.

Linear inequalities -- a system of relations among variables x_i, possibly including linear equations among them, but also including at least one inequality of the form

$$\sum a_i\ x_i \geq b$$

(in practice a strict inequality is seldom required). Such a system may be incompatible (e.g., $x_1 \geq 0$, $x_2 \geq 0$, $-x_1 - x_2 - 1 \geq 0$), may define a unique point (e.g., $x_1 \geq 0$, $x_2 \geq 0$, $-x_1 - x_2 \geq 0$), or else will define a region in space, not necessarily bounded (e.g., $x_1 \geq 0$, $x_2 \geq 0$, $-x_1 - x_2 + 1 \geq 0$ define a bounded region; $x_1 \geq 0$, $x_2 \geq 0$, $x_1 + x_2 - 1 \geq 0$ an unbounded region).

The inequality written above can be replaced by the equivalent pair

$$x_0 + \sum a_i\ x_i = b_i, \qquad x_0 \geq 0,$$

and in general it is possible to replace a system of inequalities by a system in the special form

$$A\ x = b, \qquad x \geq 0,$$

where A is a rectangular matrix, x and b are vectors.

The principal applications are in linear programming.

Linear programming -- the problem of minimizing or maximizing a linear function $C = c_0 + \sum c_i x_i$, subject to constraints in the form of linear inequalities to be satisfied by the x_i. The function C may represent costs or profits or losses in a business activity or a competition (possibly against nature), and the constraints represent limitations upon resources or their availability.

If the inequalities are compatible and satisfied by more than a single point x_i, the solution lies on the boundary, and, if unique, is the intersection of some subset of the planes whose equations are formed by using equality instead of inequality signs. Hence, in principle, the problem could be solved by solving the equations in all possible combinations and comparing the values of C. However, the difficulty of the problem comes from the fact that if the numbers of variables and relations are at all large, as they often are in practice, the number of sets to be solved may be astronomical. For such cases the simplex method may be used.

Lipschitz condition -- a condition somewhat stronger than simple continuity. In the simplest form, a function f(x) is said to satisfy a Lipschitz condition on the interval from a to b in case there exists a constant k such that

$$|f(x_1) - f(x_2)| \leq k\ |x_1 - x_2|$$

whenever x_1 and x_2 both lie on the interval. If the derivative exists and is bounded ~~continuous~~ on the closed interval then the Lipschitz condition is satisfied, but not conversely. The condition is a useful one in the study of the solution of a differential equation. Various generalizations are possible, in particular to functions of several variables.

Lobačevskiĭ method -- in current Russian literature, any root-squaring method as a method of solving algebraic equations numerically. Actually the idea of root-squaring for this purpose had already been proposed by Dandelin, and the specific algorithm proposed by Lobačevskiĭ was slightly more cumbersome than Graeffe's method which is the one in common use at present.

Lozenge diagram -- a mnemonic scheme from which can be read off a great variety of forms of the interpolating polynomial for equal-interval interpolation (see Ku for generalization). For forward difference formulas write

$$
\begin{array}{ccccccc}
f_{-2} & \binom{u+2}{1} & & & & & \\
 & & \Delta f_{-2} & & \binom{u+2}{2} & & \\
f_{-1} & \binom{u+1}{1} & & & & \Delta^2 f_{-2} & \binom{u+2}{3} \\
 & & \Delta f_{-1} & & \binom{u+1}{2} & & \\
f_{0} & \binom{u}{1} & & & & \Delta^2 f_{-1} & \binom{u+1}{3} \\
 & & \Delta f_{0} & & \binom{u}{2} & & \\
f_{1} & \binom{u-1}{1} & & & & \Delta^2 f_{0} & \binom{u}{3} \\
 & & \Delta f_{1} & & \binom{u-1}{2} & & \\
f_{2} & \binom{u-2}{1} & & & & &
\end{array}
\qquad
\begin{array}{c}
\Delta^3 f_{-2} \\
\\
\Delta^3 f_{-1}
\end{array}
$$

where the table can be extended upwards, downwards, and to the right in an obvious way. A valid polynomial for interpolation can be constructed by starting with any f_i in the first column, multiplying the binomial coefficient on the right by either of the two Δf_i to which it is joined, multiplying the coefficient on the right of that by either $\Delta^2 f_i$, continuing thus and adding results. Moreover, all polynomials

that end with the same $\Delta^n f_i$ on the right are identically equal.

The same diagram can be used for central differences if, for example, one replaces Δf_{-2} by $\delta f_{-3/2}$, $\Delta^2 f_{-1}$ by $\delta^2 f_{-1}$, ...

<u>L R transformation</u> -- a method for obtaining the <u>proper values and vectors</u> of a <u>matrix</u> A, is made by taking $A_0 = A$ and forming

$$A_i = L_i R_i, \qquad A_{i+1} = R_i L_i,$$

where R_i is upper triangular and L_i unit lower triangular. Under rather general circumstances the sequence of matrices A_i approaches an upper triangular matrix in the limit, and since all A_i have the same proper values as has A itself, the diagonal elements of the limit matrix are these proper values. The method is due to Rutishauser, who has also devised methods of accelerating convergence and for treating exceptional cases. For a discussion of the factorization see <u>matrix inversion</u>. (See NBS).

<u>Maclaurin series</u> -- <u>Taylor series</u> about the origin:

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \ldots$$

<u>Matrix</u> -- a rectangular array of scalars (real or complex) called elements of the matrix, together with certain rules of combination. If

$$A = (\alpha_{ij}) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \qquad B = (\beta_{ij})$$

and if A and B have the same dimensions, then the sum is

$$A + B = (\alpha_{ij} + \beta_{ij}),$$

that is, corresponding elements are added together.

If A has n columns and B has n rows, the product is

$$A\,B = \left(\sum_k \alpha_{ik}\,\beta_{kj}\right).$$

If ρ is a scalar,

$$\rho\,A = (\rho\,\alpha_{ij}).$$

The transpose of A is

$$A^T = (\alpha_{ji}) = \begin{pmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} & \cdots \\ \alpha_{12} & \alpha_{22} & \alpha_{32} & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix};$$

the conjugate transpose A^* is the result of replacing each element in A^T by its complex conjugate. If $A^T = A$, A is symmetric; if $A^* = A$, A is hermitian. The square matrix $I = (\delta_{ij})$, where δ_{ij} is the <u>Kronecker delta</u>, is called the identity. If it is n x n, then for any matrix A with n rows, $I\,A = A$, and for any matrix B with n columns, $B\,I = B$. A square matrix, with n rows and n columns, is said to be of order n. The square matrix C such that

$$C\,A = A\,C = I,$$

if such exists, is called the inverse of A and denoted $C = A^{-1}$. When A^{-1} exists then A is said to be nonsingular. If A is hermitian, of order n, and for every vector x of n dimensions, which is to say, matrix of n rows and one column,

$$x^*\,A\,x \geq 0,$$

then A is non-negative semidefinite; if for every vector $x \neq 0$,

$$x^*\,A\,x > 0,$$

then A is positive definite. For non-hermitian matrices these notions are not defined.

The elements α_{ii} of a matrix $A = (\alpha_{ij})$ constitute the diagonal, or main diagonal. If all other elements are null, A is a diagonal matrix. If, in addition,

$$\alpha_{ii} = \alpha,$$

that is, all diagonal elements are equal, then A is a scalar matrix and $A = \alpha I$.

If $V^* V = V V^* = I$, then V is a unitary matrix, and if V is unitary and real it is orthogonal.

If A is square of order n, then det A will represent the determinant of A. The polynomial

$$\varphi(\lambda) = \det (\lambda I - A)$$

is a polynomial in λ of degree n whose leading coefficient is unity, and whose constant term is det A. This is called the characteristic polynomial, and the Cayley-Hamilton theorem states that

$$\varphi(A) = 0.$$

There is a polynomial $\psi(\lambda)$ of minimal degree and leading coefficient unity for which

$$\psi(A) = 0.$$

This is called the minimal polynomial, and $\psi(\lambda)$ may or may not be the same as $\varphi(\lambda)$, but in all cases $\psi(\lambda)$ divides $\varphi(\lambda)$, and every zero of φ is also a zero of ψ. These zeros are variously called proper values, characteristic values, eigenvalues, latent roots. If λ is any proper value, there exists a non-null vector x satisfying

$$(A - \lambda I)\, x = 0,$$

and any such vector is called a proper vector, or characteristic vector, or eigenvector, or modal column. Naturally there exists also a non-null vector y such that

$$y^*(A - \lambda I) = 0.$$

Either x or y is said to belong to λ. If A is hermitian, then all proper values

are real. In this event there exists a unitary matrix V such that

$$V^* A V = \Lambda,$$

where Λ is diagonal, and its elements are the proper values of A. Moreover, the columns of V are proper vectors, and, in fact, if v_i is the ith volumn of V, and λ_i the ith element in the diagonal of Λ, then v_i belongs to λ_i. The matrix V is said to diagonalize A.

If $A^* A = A A^*$, then A is said to be a normal matrix. Such a matrix can also be diagonalized by a unitary matrix V, but the diagonal form Λ is not real unless A is hermitian.

For a more general matrix A, there exists a nonsingular matrix V such that

$$V^{-1} A V = \Lambda,$$

where Λ has the form

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & 0 & \cdots \\ 0 & \Lambda_2 & 0 & \cdots \\ 0 & 0 & \Lambda_3 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

and each Λ_i is either a scalar matrix or of the form

$$\Lambda_i = \begin{pmatrix} \lambda_i & 0 & 0 & \cdots \\ 1 & \lambda_i & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

with all diagonal elements equal, with ones just below the diagonal, and zeros elsewhere. This Λ is the Jordan normal form. Each λ_i is a proper value, each column v of V is a <u>principal vector</u>, which is to say that for some λ and some ν,

$$(A - \lambda I)^{\nu} v = 0.$$

Also each row of V^{-1} is a principal row vector.

For computational methods see <u>proper values and vectors</u> and <u>matrix inversion</u>.

Matrix inversion -- for a given square matrix A, the determination of a matrix denoted A^{-1} such that $A\,A^{-1} = A^{-1}\,A = I$, the identity, assuming A^{-1} to exist. A closed method of inversion would yield the exact inverse, except for errors due to rounding, after a finite number of operations in prescribed sequence performed upon the elements. These operations are usually arithmetic, but may include square-rooting. Other methods, sometimes called iterative are such as to produce a sequence of matrices C_ν which approach A^{-1} in the limit, again apart from rounding errors. Since $x = A^{-1}\,h$ satisfies the system of linear equations $A\,x = h$, and since, conversely, each column of A^{-1} is the solution of such a system with a particular h, the problems of inversion and solution are mathematically equivalent and will be spoken of interchangeably.

The most common direct methods are known as (Gaussian) elimination, or triangularization, with variants due to Crout, Banachiewicz, Doolittle, and many others, differing only in detailed arrangements and systems of recording intermediate results.

Let A be partitioned in any way in the form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} is required only to be square and nonsingular. In particular A_{11} may be a scalar, A_{21} therefore a column vector, A_{12} a row vector; or A_{22} may be a scalar, A_{12} a column vector, and A_{21} a row vector. The methods seek a factorization in one of the two forms:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}, \tag{1}$$

$$\begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}, \tag{2}$$

where the partitioning is conformable throughout. The first factorization requires that

$$L_{11} R_{11} = A_{11}, \qquad L_{11} R_{12} = A_{12},$$

$$L_{21} R_{11} = A_{21}, \qquad L_{21} R_{12} + L_{22} R_{22} = A_{22},$$

and the second that

$$P_{11} A_{11} = R_{11}, \qquad P_{11} A_{12} = R_{12},$$

$$P_{21} A_{11} + P_{22} A_{21} = 0, \quad P_{21} A_{12} + P_{22} A_{22} = R_{22}.$$

The second form is more usual but in fact they are entirely equivalent computationally. If A_{11}^{-1} is known, which is certainly the case if A_{11} is a scalar, then all matrices can be determined and one can even choose arbitrarily either P_{11} or R_{11}, and either P_{22} or R_{22}. One common choice (Gaussian elimination) is to take P_{11} the identity (= 1 if a scalar), and P_{22} the identity. Then

$$R_{11} = A_{11}, \qquad R_{12} = A_{12},$$

$$P_{21} = - A_{21} A_{11}^{-1}, \qquad R_{22} = A_{22} - A_{21} A_{11}^{-1} A_{12}.$$

Thereafter one proceeds with R_{22} as with the original A, that is, one multiplies on the left by a matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & P' \end{pmatrix},$$

where P' and R_{22} are to be partitioned. The result is the formation of n - 1 matrices P_i each unit lower <u>triangular</u>, such that

$$P_{n-1} \; P_{n-2} \; \cdots \; P_1 = P$$

is also unit lower triangular, and

$$P A = R$$

is upper triangular and hence readily inverted. Note that P has a unit determinant,

whence the determinant of A is given by the product of the diagonal elements of R. If one is interested only in solving equations, the P_i need not be retained, but each must be multiplied by the vector on the right.

While ordinarily one partitions off a row and a column, taking A_{11} a scalar, it sometimes happens that the matrix A has a submatrix A_{11} of order greater than 1, but whose inverse is already known or readily obtained. The above formulas apply equally to this case. Analogous formulas can be written for taking R_{11} and R_{22} to be identities.

In the methods of <u>enlargement</u> one applies the above formulas first to a second order submatrix; next, this is taken to be A_{11} and bordered to form a third order submatrix to be inverted by the same formulas. Eventually the entire matrix is inverted. This method is to be recommended when the matrix to be inverted is a finite segment of an infinite matrix and the size of the segment to be inverted is not determined in advance.

When A is positive definite the factorization (1) can be modified:

$$\begin{pmatrix} A_{11} & A_{12}^T \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22} \end{pmatrix},$$

with L_{11} and L_{22} both symmetric. Evidently

$$L_{11} L_{11}^T = A_{11}, \qquad L_{21} L_{11}^T = A_{21},$$

$$L_{22} = A_{22} - L_{21} L_{21}^T .$$

Ordinarily A_{11} is a scalar. For obvious reasons this is called the square-root method. It is due to Cholesky, and to Banachiewicz.

Other factorizations are to apply a Schmidt <u>orthogonalization</u> to the columns, $A = Q\,T$, where T is a unit upper triangle and Q is orthogonal by columns.

Hence $Q^T Q = D$, a diagonal, and

$$A^{-1} = T^{-1} D^{-1} Q^T.$$

It might seem more natural to form $A U = Q$ with $U = T^{-1}$ formed directly. This can be done but it turns out the same computations are required to form U directly as to form first T and then T^{-1}.

Alternatively the rows of A can be orthogonalized by applying the same process to A^T. Still another possibility is to obtain an orthogonal matrix Ω such that $\Omega A = S$, where S is upper triangular. This is done by applying first a rotation in the (1, 2)-plane to eliminate the element α_{21} of A; then in the (1, 3)-plane to eliminate α_{31}, and continuing until all subdiagonal elements are eliminated.

In the triangular factorization significance can be optimized by interchanging rows and columns at each stage so that the divisor in every division has as many significant figures as possible, or, when there is a choice, is as large as possible. In forming plane rotations, this interchange is unnecessary, but the method requires more arithmetic operations than triangular factorization, and requires a square root in the selection of each rotation.

Quite different in appearance is the method of modification based upon the following identity:

$$(B - \sigma u v^T)^{-1} = B^{-1} - \tau B^{-1} u v^T B^{-1},$$

$$\sigma^{-1} + \tau^{-1} = v^T B^{-1} u,$$

readily verified, where u and v are column vectors. If u and v are columns e_i and e_j of I, then $\sigma u v^T$ is a matrix whose only non-null element is σ in the ith row and jth column. Hence the formula gives the effect on the inverse of modifying a single element of the matrix. Or if $\sigma = 1$, $u = e_i$, it gives the result

of modifying the entire ith row by v. Evidently one could start with any matrix, perhaps I or the diagonal of A, whose inverse is known, and modify row by row, or column by column, to build up to A^{-1}. In case A is symmetric, take $u = v = e_i \pm e_j$, $\sigma = \alpha_{ij} = \alpha_{ji}$. This affects also the diagonal elements so that one starts by inverting a suitably demodified diagonal of A. This method seems to be in principle the method of "tearing" advocated by Kron.

If C is an approximation to A^{-1} which one wishes to improve, form either of

$$H = I - A\,C, \qquad K = I - C\,A.$$

Then either of

$$C_1 = C(I + H), \qquad C_1' = (I + K)\,C$$

may be a better approximation, and will be if in any <u>norm</u>, one of the following is true:

$$\|H\| < 1, \qquad \|K\| < 1.$$

Convergence is then quadratic. A rigorous bound for the error in the inverse is given by either of

$$\|A^{-1} - C\| \le \|C\,H\| \,/\, (1 - \|H\|),$$

$$\|A^{-1} - C\| \le \|K\,C\| \,/\, (1 - \|K\|),$$

under the same condition.

For solving the system

$$A\,x = h,$$

if A is of high order but with most elements null, one usually prefers to generate a sequence of vectors x_ν approaching the solution x in the limit. Let

$$S_\nu = x - x_\nu, \qquad r_\nu = h - A\,x_\nu = A\,S_\nu.$$

Starting with any x_0, one forms recursively

$$x_{\nu+1} = x_\nu + C_\nu r_\nu,$$

where C_ν is some matrix which may be fixed in advance, or may change from step to step. One can always arrange, if convenient, so that every diagonal element of A is unity. If this is done, and $C_\nu = I$, one has the simple, or total step, iteration. If one writes, then,

$$A = I - B,$$

so that B has a null diagonal, it is necessary and sufficient that the <u>spectral radius</u> $\rho(B) < 1$. For the single-step (or Seidel, or Gauss-Seidel, or Liebmann) iteration let

$$A = L + R,$$

where L is null except below the diagonal, and R is null below the diagonal. Then

$$C_\nu = R^{-1},$$

or, as the computation is usually made, one solves

$$R\, x_{\nu+1} = h - L\, x_\nu.$$

Convergence is assured when A is positive definite, and, more generally, when $\rho(R^{-1} L) < 1$, the latter condition being necessary and sufficient. Note that in solving for the ith element of $x_{\nu+1}$, one solves the ith equation for that element in terms of the most recent values of all other elements, and the equations are taken in fixed cyclic order. In the method of <u>relaxation</u> there is no fixed order, but one selects at each stage the particular equation which is least well satisfied, and if this is the jth, one solves for the jth element in terms of most recent values of the others. This inspection can be done readily in computing by hand, but not readily by machine, whence relaxation is not recommended for machine computation (although some authors apply the term more generally to any method, or at least some other methods, of generating converging sequences). In the method

of steepest descent

$$C_\nu = (r_\nu^T A\, r_\nu)^{-1}\, r_\nu\, r_\nu^T.$$

The method is more usually written

$$x_{\nu+1} = x_\nu = \mu_\nu\, r_\nu,$$

$$\mu_\nu = r_\nu^T r_\nu \,/\, r_\nu^T A\, r_\nu.$$

It applies in this form only when A is positive definite, and then it always converges. Note, however, that any system with a nonsingular matrix A can be converted to an equivalent system with a positive definite matrix:

$$A^T A\, x = A^T h.$$

See condition number, norm.

Memory unit (of a computer) -- see storage unit.

Method of successive approximation -- see iterative methods.*

Milne method -- for solving an ordinary differential equation

$$f' = f(x, y)$$

uses the predictor

$$y_{\nu+1} = y_{\nu-3} + 4h(2f_\nu - f_{\nu-1} + 2f_{\nu-2})$$

and the corrector

$$y_{\nu+1} = y_{\nu-1} + h(f_{\nu+1} + 4f_\nu + f_{\nu-1}) \,/\, 3,$$

the latter formula being the Simpson rule applied to the evaluation of the integral. The formulas are meaningful when y and f are vectors. The method is one of the simpler ones and often quite accurate, but for some problems exhibits serious lack of stability (see Mi, Co).

*

Midpoint formula -- for numerical quadrature,

$$\int_a^b f(x)\, dx = (b-a)\left(\frac{a+b}{2}\right) + \frac{1}{24}(b-a)^3 f''(\xi),$$

$$a \le \xi \le b.$$

Minimax approximation -- the representation of a function $f(x)$ by an approximation $P[f \mid x]$ selected from a class of functions in such a way that if $R = f - P$, then the maximum of $|R|$ over all points x of some assigned set is minimized over all functions of the class. Often called Chebyshev approximation. There is no simple algorithm for forming the Chebyshev approximation to an arbitrary continuous function, but there are techniques of successive approximation to it.

In case the class of functions is the class of linear combinations of functions $\varphi_0(x)$, $\varphi_1(x)$, ..., $\varphi_n(x)$ on an interval from a to b, then the basic theorem is the following: If the functions φ_i form a Chebyshev system of order n for the interval from a to b, and if f is continuous on this interval, then the minimax approximation $P_n(x)$ is unique; moreover $|f(x) - P_n(x)|$ achieves its maximum at least $n + 2$ times at points x_0, x_1, ..., x_{n+1} on the interval and at these points $f - P_n$ alternates in sign.

The Chebyshev polynomial $T_n(x)$ is that polynomial of degree n with leading coefficient unity that gives the minimax approximation to zero on the interval from -1 to $+1$; hence $x^n - T_n$ is that polynomial of degree $n - 1$ that gives the minimax approximation to x^n. This explains the utility of Chebyshev expansions (see Ko, La, Ha).

Modal column -- of a matrix, a term sometimes applied to a proper vector. See proper values and vectors.

Monte Carlo method -- any method for obtaining a statistical estimate of a desired quantity by random sampling. In the most successful applications the desired quantity is in fact a statistical parameter, and the sampling is made from an artificial population that is in some sense a model of the physical system itself: for example, the fraction of neutrons, of given initial distribution in direction and energy, that can be expected to penetrate to a given depth in a medium of known composition and geometric form. For more details see Ho, MC, Sym.

Multivariate interpolation -- interpolation of functions of two or more variables. Since

$$E_x = 1 + \Delta_x, \qquad E_y = 1 + \Delta_y$$

one can expand formally

$$E_x^u E_y^v = (1 + \Delta_x)^u (1 + \Delta_y)^v$$

and derive equal interval formulas. Likewise analogues of the Lagrange formula can be derived. To write down explicitly is hardly worth while since they amount to operating on the two (or more) variables separately (see Ku, W and R).

Neumann series -- the series

$$\varphi(x) = f(x) + \lambda \int K(x, y) f(y) d y + \lambda^2 \iint K(x, y) K(y, z) f(z) d y d z + \ldots,$$

with appropriate limits of integration. If the series converges, it converges to a solution of the integral equation

$$\varphi(x) = f(x) + \lambda \int K(x, y) \varphi(y) d y.$$

Neville method of interpolation -- see Aitken method of interpolation.

Newton-Cotes quadrature formulas -- these are of the form

$$\int_{x_0}^{x_n} f(x) d x = \eta_0 f_0 + \eta_1 f_1 + \ldots + \eta_n f_n + R_n,$$

where the coefficients η_i are obtained by integrating the coefficients of the Lagrange interpolation formula. The best known and most useful of these are the trapezoidal rule, Simpson's rule, and the three-eighths rule.

Newton-Gauss interpolation formulas -- see Gaussian interpolation formulas.

Newton interpolation formulas -- see difference operators.

Newton-Stirling interpolation formulas -- see Stirling interpolation formulas.

Newton method -- for solving equations, whether algebraic or transcendental, is a particular iterative method, and the one most commonly used. The method can also be extended to simultaneous equations and to matrix inversion.

Nomograph or nomogram -- also called an alignment chart or isopleth. It consists of two or more scales, drawn and arranged so that results of calculation may be found from the relation of points on them. For example, suppose the relation $f(x, y, z) = 0$ is given. Three graduated scales or lines are then constructed so that, if selected values of two variables are located with a straight-edge, the third scale will be cut at a value which satisfies the given equation. The method is readily extended to more than three variables.

Norm -- in some literature the norm $\|x\|$ of the vector x, and the norm $\|A\|$ of the matrix A refers specifically to what will be called the Euclidean norm, defined by the non-negative square roots of

$$\|x\|^2 = x^* x \; ; \qquad \|A\|^2 = \text{trace } (A^* A).$$

(See trace.) More generally, however, a norm $\|x\|$ of a vector x is taken to be any real valued function of the elements satisfying the conditions

i) $x \neq 0 \Longrightarrow \|x\| > 0;$

ii) $\|\alpha x\| = |\alpha| \; \|x\|;$

iii) $\|x + y\| \leq \|x\| + \|y\|;$

and the associated norm $\|A\|$ of a square matrix A is

$$\|A\| = \sup_{\|x\| = 1} \|A x\|.$$

Such a matrix norm possesses properties (i), (ii), and (iii), with matrix arguments, and also, for any two matrices A and B,

iv) $\|A B\| \leq \|A\| \; \|B\|.$

The Euclidean matrix norm possesses these properties, but does not satisfy the definition. The Euclidean matrix and vector norms are further related, as are all associated norms, by the property that for any matrix A and vector x,

$$\|A\,x\| \le \|A\|\ \|x\|.$$

But for norms as here defined, for any A there exists an $x \neq 0$ such that an equality holds, and this is not true for the Euclidean norms. In fact, it is the <u>spectral norm</u> that is associated with the Euclidean vector norm.

Norms are important in studying convergence properties of sequences of vectors and matrices, and in obtaining error bounds (see <u>matrix inversion</u>). In fact, a sequence of vectors x_ν has a limit x if and only if the sequence of norms $\|x_\nu - x\|$ vanishes in the limit, whatever norm may be used. A sufficient condition for the sequence of matrix powers B^ν to have the limit 0 is that $\|B\| < 1$, in any norm. Also for any matrix A and any norm, the <u>spectral radius</u> satisfies

$$\rho(A) \le \|A\|.$$

If x satisfies $A\,x = h$, if $y = x - s$ is any approximation, and if $r = h - A\,y = A\,s$, then

$$\|s\| \le \|A^{-1}\|\ \|r\|,$$

which gives a measure of the departure of the approximate solution y from the true solution x. In some cases $\|A^{-1}\|$ can be evaluated a priori, as when the equations arise from a partial differential equation. Otherwise if $H = I - A\,C$, and $\|H\| < 1$, then

$$\|A^{-1}\| \le \|C\| \,/\, (1 - \|H\|).$$

As other examples of norms, if $x^T = (\xi_1, \ldots, \xi_n)$ and $A = (\alpha_{ij})$, then

$$\|x\|_e = \max_i |\xi_i|\ , \qquad \|A\|_e = \max_i \sum_j |\alpha_{ij}|\ ;$$

$$\|x\|_{e'} = \sum_i |\xi_i|\ , \qquad \|A\|_{e'} = \max_i \sum_j |\alpha_{ji}|\ .$$

Again, given any norm, and any nonsingular matrix G (for example, a diagonal matrix),

$$\|x\|_G = \|G^{-1} x\| \; ; \qquad \|A\|_G = \|G^{-1} A G\|$$

are again norms.

Numerical analysis -- the theory underlying the development of numerical processes. In Lanczos's terms, numerical analysis is distinguished from parexic analysis which has as its aim the development of approximate formulas. In more common usage it would include parexic analysis.

Numerical cubature -- see multivariate interpolation.

Numerical differentiation -- see differentiation formulas.

Numerical quadrature -- see quadrature.

Numerical solution of differential equations -- see differential equations.

Numerical solution of equations -- see algebraic equations, matrix inversion, Newton method, iterative methods, Bernoulli method, Graeffe method, Horner method.

Numerical stability -- A problem is numerically stable in case small errors in the data produce moderate errors at most in the required quantities. A method is numerically stable if it does not permit the accumulation of large generated errors, or if it does not involve a transformation of the original problem into another one that may be substantially less stable. The term is applied in particular to certain methods for solving ordinary and partial differential equations for which the approximating system of difference equations is stable. (See condition number.)

Open quadrature formula -- a quadrature formula in terms of ordinates drawn at points within the interval of integration, and excluding at least one endpoint, e.g., for integrating from x_0 to x_n, but based upon $f_0, f_1, \ldots, f_{n-1}$ only and not involving f_n. Such a formula is useful as a predictor formula in the numerical solution of ordinary differential equations.

Operational methods -- the treatment of differential and other operators as though they were algebraic symbols in order to derive formal identities. Most often the term applies to the use of Fourier, Laplace, or other transforms in the solution of functional equations, but see difference operators for the formal derivation of formulas for interpolation, quadrature, and differentiation.

Optimum-interval interpolation -- the use of interpolation formulas based upon intervals of varying size adjusted so as to optimize the result in some way. Usually the interval size will be taken as great as possible so that the error of an interpolating polynomial of fixed degree will not surpass a specified limit (see Ho).

Ordinary differential equation -- one in which there occur, along with each dependent variable, one or more of its derivatives with respect to a single independent variable. In general there must be as many equations as they are dependent variables. Also, in general, if any derivatives higher than the first occur, it is possible to enlarge the system to one containing only first derivatives. Thus if the equation is of the form

$$f(x, y, y', y'') = 0,$$

where x is the independent variable, one can set

$$y_1 = y, \qquad y_2 = y',$$

and the given equation is equivalent to the simultaneous equations

$$f(x, y_1, y_2, y_2') = 0,$$
$$y_1' - y_2 = 0.$$

Finally, it is generally possible to solve for first derivatives and write the system in the form

$$d\, y_i \,/\, d\, x = \varphi_i(x, y_1, y_2, \ldots, y_n), \quad i = 1, 2, \ldots, n.$$

The equations are then said to be in normal form. In order that the $y_i(x)$ be uniquely defined as functions of x it is necessary to adjoin n further conditions relating the values assumed by the y_i at certain discrete points. For an initial value problem, all values are specified at a single point; for a <u>boundary-value problem</u>, relations between values at the two boundaries are specified. Here it will be assumed that initial values are given.

A system in normal form can be compactly represented by

$$y' = f(x, y),$$

where, if $n > 1$, y, y', and f represent vectors in n-space. Methods for numerical solution in the general case are rather obvious generalizations of the case $n = 1$, and attention will be confined to this. It will be assumed that the initial value

$$y(x_0) = y_0$$

is given. In case $f(x, y)$ ~~possesses partial derivatives of all orders~~ is analytic in x and y at (x_0, y_0), then it is possible to obtain the coefficients of the series expansion

$$y = y_0 + (x - x_0)\, y_0' + \tfrac{1}{2}(x - x_0)^2\, y_0'' + \ldots.$$

In fact, y_0 is ~~known~~ given, from the equation itself one obtains

$$y_0' = f(x_0, y_0),$$

and by differentiation one obtains

$$y_0'' = f_x(x_0, y_0) + y_0'\, f_y(x_0, y_0) \quad .$$

and other derivatives of higher order are similarly found.

Most numerical methods, however, combine interpolation with the <u>Picard method</u> in the sense that the quadratures required by this method are carried out numerically. This idea is capable of almost infinitely many variations, but to begin with it is generally assumed that by series expansion or otherwise one has first obtained suitable values $y_1 = y(x_1)$, $y_2 = y(x_2)$, ... where, in general,

$x_\nu = x_0 + \nu h$, and h represents a sufficiently fine subdivision. Thereafter one can write

$$y_{\nu+i} = y_\nu + \int_{x_\nu}^{x_{\nu+i}} f(x, y)\, d\, x,$$

where $y_\nu, y_{\nu+1}, \ldots, y_{\nu+i-1}$ have been found. At this point one uses first a predictor formula which expresses the integral in terms of values of f at points already known (see quadrature, numerical). This gives an approximate value for $y_{\nu+i}$, hence permits the calculation of an approximate value of $f_{\nu+i}$, and one can then apply a corrector formula which utilizes this value. Application of the corrector formula gives a new approximation for $y_{\nu+i}$, and the corrector formula should be reapplied until the changes are negligible. It is advisable to have h small enough so that at most a second application of the corrector formula will be sufficient. Methods of this type are the Euler method, the Adams-Bashford method, and the Milne method (see NPL, Mi, Co). For questions of stability see Mi, Co.

The method of Runge-Kutta is somewhat different in character.

For the explicit treatment of an equation of higher order

$$y'' = f(x, y, y')$$

without reducing it to normal form in two independent variables, see Störmer method.

Orthogonalization -- Given an inner product operation (φ, ψ), defined for any pair of functions φ and ψ of a specified class, if $(\varphi, \psi) = 0$ then φ and ψ are said to be orthogonal. Given a sequence of functions of the class,

$$\varphi_0, \varphi_1, \varphi_2, \cdots$$

one can ask for a sequence

$$\psi_0, \psi_1, \psi_2, \cdots$$

such that $(\psi_i, \psi_j) = 0$ when $i \neq j$, and

$$\varphi_0 = \psi_0,$$
$$\varphi_1 = \alpha_{10}\,\psi_0 + \psi_1,$$
$$\varphi_2 = \alpha_{20}\,\psi_0 + \alpha_{21}\,\psi_1 + \psi_2,$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

where the α_{ij} are constants. One finds easily that

$$(\psi_0, \varphi_1) = \alpha_{10}\,(\psi_0, \psi_0),$$
$$(\psi_0, \varphi_2) = \alpha_{20}\,(\psi_0, \psi_0),$$
$$(\psi_1, \varphi_2) = \alpha_{21}\,(\psi_1, \psi_1),$$

and all α_{ij} and ψ_i can be obtained sequentially unless for some i, $(\psi_i, \psi_i) = 0$.

Orthogonal matrix -- see matrix.

Orthogonal polynomials -- any set of polynomials that are mutually orthogonal with respect to an inner product operation. Such an operation can be defined quite formally in terms of an abstract moment operator Ω defined by

$$\Omega(\alpha_0 + \alpha_1\,x + \alpha_2\,x^2 + \ldots) = \alpha_0\,\mu_0 + \alpha_1\,\mu_1 + \alpha_2\,\mu_2 + \ldots$$

where the μ_i are arbitrarily assigned constants. Most often one defines the result of Ω operating on a function φ to be

$$\Omega[\varphi(x)] = \int_a^b \varphi(x)\,\omega(x)\,d\,x$$

where a and b are fixed, finite or infinite, and $\omega(x)$ is a fixed ~~weight~~ density function. The inner product of two polynomials φ and ψ is

$$(\varphi, \psi) = \Omega\,[\varphi(x)\,\psi(x)].$$

The orthogonal polynomials can be obtained by applying the process of orthogonalization to the powers of x:

$$1,\ x,\ x^2,\ x^3,\ \ldots.$$

Orthogonal polynomials arise in Gaussian quadrature in that their zeros are the points x_i at which the integrand function must be evaluated. They occur as

the denominators of the entries in a Padé table along a line parallel to the diagonal.

Given a finite set of discrete points x_0, x_1, ..., if the operator Ω is defined by

$$\Omega[\varphi(x)] = \sum \varphi(x_i)\, \omega_i$$

where ω_i are non-negative weight functions, then the corresponding orthogonal polynomials are finite in number. Let these be ψ_0, ψ_1, ψ_2, ... For least square polynomial fitting to data taken at these points x_i, it is advantageous to express the required polynomial in the form

$$f(x) = \gamma_0\, \psi_0 + \gamma_1\, \psi_1 + \gamma_2\, \psi_2 + \cdots$$

and to solve for the γ's by least squares.

Overrelaxation -- For solving the system

$$x = k + B\, x$$

(see matrix inversion) by the simple iteration scheme one writes

$$x_{\nu+1} = x_\nu + r_\nu, \quad r_\nu = k - (I - B)\, x_\nu.$$

In case of slow convergence it is suggested to write $\alpha_\nu\, r_\nu$ in place of r_ν, where α_ν is a scalar. When $\alpha_\nu > 1$ this is known as overrelaxation, when $\alpha_\nu < 1$, underrelaxation. When $\rho(B) < 1$ (which is necessary for convergence) and all proper values are real, an optimal choice for α_ν can be found by the use of Chebyshev polynomials.

Padé table -- a double-entry table associated with a formal power series

$$f(z) = s_0 + s_1\, z + s_2\, z^2 + \cdots.$$

The entry in row p and column q, where p, q ≥ 0, is the fraction $A_{p,q}(z) \,/\, B_{p,q}(z)$, where $A_{p,q}$ and $B_{p,q}$ are polynomials of degree q and p, respectively, such that the formal expansion of the fraction as a power series in z agrees with that of f to

and including the term in z^{p+q}. The consecutive entries taken in a diagonal direction are approximants of a continued fraction expansion of $f(z)$. Often the continued fraction will converge more rapidly than the power series, or may even converge where the power series diverges, and so provides an effective computational device. To form a particular entry in the Padé table, let

$$B_{p,\,q} = b_0 + b_1 z + \ldots + b_p z^p,$$

and observe that in the product $f(z)\, B_{p,q}(z)$, terms in z^{q+1}, z^{q+2}, ..., z^{q+p} all drop out. This provides p homogeneous equations in the p + 1 unknowns, b_0, ..., b_p. If a solution exists, then the polynomial $A_{p,q}$ can be obtained by direct multiplication. See quotient-difference algorithm, orthogonal polynomials. The Shanks generalizations of the delta-square process applied to the sequence of partial sums of a power series lead to entries in the Padé table.

Parabolic equation -- a partial differential equation that can be reduced to the form

$$\partial^2 u / \partial x^2 = \partial u / \partial t$$

when the variables are suitably chosen.

Parexic analysis -- a term coined by Cornelius Lanczos to designate the branch of analysis concerned with the derivation of "nearby" (i.e., approximating) but numerically manageable relations. It is distinguished, on the one hand, from "pure" analysis, which is concerned with limiting processes and exact relations. On the other hand it is distinguished from numerical analysis which is concerned with the translation into numerical processes of the approximations provided by parexic analysis. The aim of parexic analysis is to reduce truncation errors; that of numerical analysis to reduce generated error and the number of operations required.

Partial differential equation -- a differential equation containing two or more independent variables, hence containing partial derivatives. The order of the equation is that of a derivative of highest order. It is linear if linear in the dependent variable(s) and all derivatives; quasilinear if linear in the partial derivatives of highest order. Second-order linear differential equations

$$a\, u_{xx} + 2b\, u_{xy} + c\, u_{yy} + \ldots = 0$$

in two independent variables are classified as

elliptic,	$a\,c - b^2 > 0$,
parabolic,	$a\,c - b^2 = 0$,
hyperbolic,	$a\,c - b^2 < 0$.

The coefficients a, b, and c are functions of x and y, and the inequalities must hold throughout the region under consideration.

Methods of solution differ slightly for the three types. Most often, however, for any type and any order, a method of differences is used by which the differential operators are replaced by approximating expressions in terms of difference operators. This requires dividing the space of the independent variables, in the case of two independent variables, by equally spaced lines parallel to each axis; for three independent variables by equally spaced planes parallel to each coordinate plane; and similarly in higher dimensions.

The method is best described by an example. Consider the simple parabolic equation

$$u_{xx} = u_y,$$

or, in operator form,

$$\left(D_x^2 - D_y\right) u = 0.$$

The natural, and simplest, difference representation of D_x^2 is by central differences

$$D_x^2 \doteq h^{-2}\, \delta_x^2,$$

where h represents the constant difference along the x-axis. There are two equally

natural representations of D_y:

$$D_y \doteq k^{-1} \Delta_y, \qquad D_y \doteq k^{-1} \nabla_y.$$

If applied to $u_{ij} = u(x_i, y_j)$, these lead to somewhat different difference equations which can be written:

$$\gamma\, u_{i+1,j} + (1 - 2\gamma)\, u_{i,j} + \gamma\, u_{i-1,j} - u_{i,j+1} = 0,$$

$$\gamma = k / h^2,$$

in the first case, and, in the second case

$$\gamma\, u_{i+1,j} - (1 + 2\gamma)\, u_{i,j} + \gamma\, u_{i-1,j} + u_{i,j-1} = 0.$$

The coefficients in the two approximations can be represented schematically ~~in the~~ by the following ~~way:~~ stars:

$$\begin{array}{ccc} & -1 & \\ \gamma & 1 - 2\gamma & \gamma \end{array} \qquad\qquad \begin{array}{ccc} & & \\ \gamma & -(1 + 2\gamma) & \gamma. \\ & 1 & \end{array}$$

These schemes show the relative locations of the points to which the coefficients are applied. The first scheme is simplest to apply, since one has only to solve for u at the upper point in terms of its values at the points below. It turns out, however, that to insure adequate <u>numerical stability</u> it is necessary that $\gamma \leq 1$. This is an example of an explicit (or marching) scheme. To apply the other method no restriction is needed on γ, but to obtain the value of u along any horizontal line it is necessary to solve a system of linear equations, although the matrix is tridiagonal, and hence readily inverted. This is an example of an implicit (or jury) scheme.

To start the solution it is necessary that suitable initial and boundary conditions be given. If these involve derivatives, then the derivatives in the boundary conditions must also be represented by approximating differences.

Picard method -- a numerical method of successive approximations or iteration for solving differential equations. If the given equation is $y' = f(x, y)$ subject to the condition that $y = y_0$ when $x = x_0$, the solution may be written in the form of an integral equation

$$y = y_0 + \int_{x_0}^{x} f(x, y)\, d\, x.$$

An approximate solution is

$$y_1 = y_0 + \int_{x_0}^{x} f(x, y_0)\, d\, x,$$

and a sequence of approximations can be formed:

$$y_2 = y_0 + \int_{x_0}^{x} f(x, y_1)\, d\, x\ ;$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$y_n = y_0 + \int_{x_0}^{x} f(x, y_{n-1})\, d\, x.$$

Under quite general circumstances this sequence can be shown to converge to the solution. Unless f is a polynomial it is not generally feasible to effect the quadratures analytically, but most methods of numerical solution of differential equations take their departure from this and integrate numerically. Moreover, y and f can be interpreted as vectors and the methods apply equally to systems of differential equations.

Planimeter -- an instrument for measuring the area enclosed by a curved boundary, hence an analog computer for the purpose of performing quadratures.

Poisson equation -- see Laplace differential equation.

Positive (semi) definite matrix -- a matrix that is Hermitian with all proper values positive (non-negative).

Precision -- see accuracy.

Predictor -- a quadrature formula used in the numerical solution of ordinary dif-diferential equations in which the integral is expressed in terms of equally spaced ordinates to the left of but not including the endpoint. This provides an initial approximation to the new ordinate. Thereafter the approximation is used in a corrector to improve or check the approximation. A predictor formula is open, a corrector formula closed. See Adams-Bashford method for an example.

Principal vector -- of a matrix A, any non-null vector x such that for some scalar λ and integer ν,

$$(\lambda I - A)^{\nu} x = 0.$$

In case $\nu = 1$, x is a proper vector, and in any event λ is a proper value.

Program -- a plan or design of a computational routine to be performed by a stored-program computer. Sometimes this term is loosely used interchangeably with code.

Proper values and vectors (of matrices, methods of computing) -- For a given square matrix A (singular or not), non-null vectors x exist satisfying $A x = \lambda x$ only when the scalar λ satisfies $\det(\lambda I - A) = \varphi(\lambda) = 0$, called the characteristic equation, where $\varphi(\lambda)$ is a polynomial of degree n, the order of A, called the characteristic polynomial (see matrix). Any root λ of this equation is called a proper value; associated with any root λ there is at least one non-null vector x called a proper vector belonging to λ. The number of independent proper vectors belonging to a proper value may equal, but cannot exceed, the multiplicity of λ as a root of the characteristic equation. For normal matrices, including hermitian, the number equals the multiplicity. For a non-normal matrix, if λ is a root of multiplicity $\nu > 1$, and if fewer than ν independent vectors belong to λ, then there are principal vectors $y \neq 0$ satisfying $(\lambda I - A)^{\mu} y = 0$ for $1 < \mu \leq \nu$. It is always true that $\varphi(A) = 0$. If there is a polynomial $\psi(\lambda)$ of degree $m < n$ for which $\psi(A) = 0$, then that polynomial of lowest degree for which this is true is a minimal polynomial,

and A is said to be derogatory; otherwise nonderogatory. The condition for being derogatory is rather stringent and not often satisfied in practice, and the same is true for the presence of principal vectors, but when the conditions are nearly satisfied the corresponding proper vectors are poorly defined and computational difficulties arise, possibly even insurmountable.

Implicitly or explicitly, to evaluate a proper value requires the solution of an algebraic equation of degree n. The problem is much simpler for a hermitian matrix, and for these will be discussed under that heading. Some of these methods can also be adapted to normal matrices. Here the general case will be considered. A method may purport to yield only the characteristic, or possibly the minimal, polynomial, leaving this to be solved by any of the standard methods for solving algebraic equations. Such a method is called direct. Once a proper value is known, a proper vector belonging to it can be found by solving a system of homogeneous equations (see matrix inversion). A direct expansion of the determinant is unthinkable when n is at all large.

Since any n + 1 vectors in n-space are linearly dependent, if $b = b_1 \neq 0$, then in the sequence

$$b_{i+1} = A\, b_i$$

there is a smallest index $m \leq n$ such that b_{m+1} is linearly dependent upon those preceding:

(1) $$b_{m+1} + \beta_1\, b_m + \beta_2\, b_{m-1} + \ldots + \beta_m\, b_1 = 0.$$

This represents n equations in $m \leq n$ unknowns, and they are consistent by hypothesis, even if $m < n$. But this is

(2) $$(A^m + \beta_1\, A^{m-1} + \ldots + \beta_m\, I)\, b = 0.$$

Hence, if

(3) $$\psi(\lambda) \equiv \lambda^m + \beta_1\, \lambda^{m-1} + \ldots + \beta_m\,,$$

then $\psi(A)\, b = 0$ and $\psi(\lambda)$ is either a minimal polynomial or a factor of it, and except for very special choices of b, $\psi(\lambda)$ is a minimal polynomial.

If one forms the matrix

(4) $$B = (b_1, b_2, \ldots, b_m)$$

whose columns are the b_i, then

(5) $$A\, B = B\, F$$

where F is the companion matrix whose form is

(6) $$F = \begin{pmatrix} 0 & 0 & \ldots & 0 & -\beta_m \\ 1 & 0 & \ldots & 0 & -\beta_{m-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & -\beta_1 \end{pmatrix}.$$

Its characteristic polynomial is $\psi(\lambda)$. With $b_1 = e_1$, the first column of I, this is the method of Krylov.

This method is effective for matrices of fairly low order. For matrices of high order the equations to be solved for the β_i tend to become ill-conditioned. A method due to Hessenberg yields a matrix B such that

(7) $$A\, B = B\, G,$$

where G is not the companion matrix, but whose form is such that the characteristic polynomial can be expanded directly, and from which it is rather easy to solve for the proper vectors. To apply this method one selects auxiliary vectors $c_1, c_2, c_3, \ldots$ subject to two mild restrictions but otherwise arbitrary. Ordinarily one can take (with Hessenberg) each $c_i = e_i$, a column of I. One starts again with a vector b, possibly e_1. Now b_2 is a linear combination of b_1 and $A\, b_1$ orthogonal to c_1; b_3 a linear combination of b_1, b_2, and $A\, b_2$ orthogonal to both c_1 and c_2; ... The restrictions on the c_i are, first, linear independence, and, second,

$$c_i^*\, b_i \neq 0.$$

With these conditions fulfilled one forms

$$A\, b_1 = b_1\, \gamma_{11} + b_2$$

where

$$c_1^*\, A\, b_1 = c_1^*\, b_1\, \gamma_{11}.$$

Hence γ_{11} is obtained, and therefore b_2. Next

$$A\, b_2 = b_1\, \gamma_{12} + b_2\, \gamma_{22} + b_3,$$

where

$$c_1^*\, A\, b_2 = c_1^*\, b_1\, \gamma_{12},$$

$$c_2^*\, A\, b_2 = c_2^*\, b_1\, \gamma_{12} + c_2^*\, b_2\, \gamma_{22}.$$

From the first of these one obtains γ_{12}, from the second γ_{22}, and finally b_3. The process continues, and G is seen to have the form

$$G = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots \\ 1 & \gamma_{22} & \gamma_{23} & \cdots \\ 0 & 1 & \gamma_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{8}$$

It may be observed that when the choice $c_i = e_i$ is made, then the first element of b_2 is null as are the first two elements of b_3, the first three of $b_4, \ldots$

A further refinement of the method is the biorthogonalization method of Lanczos. In this method only b_1 and c_1 are arbitrary. Thereafter, just as b_2 is a linear combination of b_1 and $A\, b_1$ orthogonal to c_1, so is c_2 a linear combination of c_1 and $A^*\, c_1$ orthogonal to b_1, and so for the others. The matrix G is then tridiagonal, being null everywhere except along, just below, and just above the diagonal. The recursion then never contains more than three terms in any equation.

Another modification is due to Arnoldi, who chooses $b_i = c_i$ and requires B to be an orthogonal matrix. Then (7) is satisfied, but with a matrix G of the form

$$G = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \cdots \\ 0 & \gamma_{32} & \gamma_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \tag{9}$$

similar to (8) but with elements other than ones on the subdiagonal. The development of the characteristic polynomial from G is almost equally straightforward.

When only one, or few, of the proper values of largest modulus are required a simple iteration scheme is advisable, and is preferred by some even for a complete reduction of rather large matrices. If there is a single proper value exceeding all others in modulus, λ_1, vectors b_ν in the continued sequence

$$b_{\nu+1} = A\, b_\nu$$

approach the proper vector belonging to λ_1, so that for large ν,

$$b_{\nu+1} \doteq \lambda_1\, b_\nu.$$

Hence, with any vector u, an approximate value of λ_1 is given by

$$u^* A\, b_\nu = \lambda_1\, u^* b_\nu.$$

Convergence, when it occurs, can be accelerated by application of the <u>delta square process</u> termwise to the vectors in the sequence.

In the case of a pair of complex roots, convergence does not occur, but in the limit b_ν becomes parallel to the plane of the two proper vectors. Then if

$$\mu_\nu = u^* b_\nu$$

the roots of the equation

$$\begin{vmatrix} 1 & \mu_\nu & \mu_{\nu+1} \\ \lambda & \mu_{\nu+1} & \mu_{\nu+2} \\ \lambda^2 & \mu_{\nu+2} & \mu_{\nu+3} \end{vmatrix} = 0$$

approach these roots λ_1 and λ_2 as ν increases. In fact, whether or not λ_1 and λ_2 are equal in modulus, but provided only both exceed all others in modulus, the statement is true. Moreover, the roots of

$$\begin{vmatrix} 1 & \mu_\nu & \mu_{\nu+1} & \mu_{\nu+2} \\ \lambda & \mu_{\nu+1} & \mu_{\nu+2} & \mu_{\nu+3} \\ \lambda^2 & \mu_{\nu+2} & \mu_{\nu+3} & \mu_{\nu+4} \\ \lambda^3 & \mu_{\nu+3} & \mu_{\nu+4} & \mu_{\nu+5} \end{vmatrix} = 0$$

approach the three proper values of largest modulus, if such exist (cf. <u>Bernoulli method</u>).

Returning to the two roots, for large ν,

$$b_\nu \doteq v_1 + v_2,$$
$$b_{\nu+1} \doteq \lambda_1 v_1 + \lambda_2 v_2,$$

approximately, where v_1 and v_2 are proper vectors belonging to λ_1 and λ_2. Knowing λ_1 and λ_2, as well as b_ν and $b_{\nu+1}$, these equations can be solved for v_1 and v_2.

When any proper value and proper vector are known it is possible to apply "deflation" as follows: The matrix

$$A - \lambda_1 v_1 u^*,$$

where u is any vector satisfying $u^* v_1 = 1$ has the same proper values as A except that instead of λ_1 the new matrix has a zero. Moreover, if w is a proper vector of $A - \lambda_1 v_1 u^*$, belonging to the proper value λ, then $(A - \lambda_1 I) w$ is a proper vector of A belonging to the proper value λ. And finally, it is possible to choose u^* so that every element but the first in the first row of $A - \lambda_1 v_1 u^*$ is zero. To do this, normalize v_1 to have 1 as the first element (should this be zero, permute rows and corresponding columns of A, and the same elements of v_1 to make the first element non-null). Choose the last n - 1 elements of u^* to be $1/\lambda_1$ times

the corresponding elements of the first row of A, and choose the first element of u^* to satisfy $u^* v_1 = 1$. It remains now to consider only a matrix of order n - 1, and further reductions are made as further proper values and vectors are found.

In case μ is an approximation to a particular proper value λ_i, closer than to any other λ_j, then $\mu - \lambda_i$ is the numerically largest proper value of $(A - \mu I)^{-1}$. Hence the solution of the system

$$(A - \mu I)\, x_1 = x_0$$

for an almost arbitrary x_0 will give a good approximation to the proper vector belonging to λ_i, and $x_1^* A x_1 / x_1^* x_1$ will be a better approximation to λ_i. This is Wielandt's broken iteration.

Unless it is known a priori that a matrix has <u>principal vectors</u> that are not proper vectors, and has auxiliary conditions available for determining them, rounding errors will almost inevitably conceal them, but the problem may turn out to be highly unstable. If $A = P \Lambda P^{-1}$ where Λ is diagonal, the <u>numerical stability</u> decreases as $\|P\| \, \|P^{-1}\|$ increases in some <u>norm</u>, hence this seems to be an appropriate <u>condition number</u> with respect to this problem. But as A approaches the nondiagonalizable form, this number becomes infinite, hence at least some of the proper values and vectors become ill defined.

<u>Proper vector</u> -- of a <u>matrix</u>, see <u>principal vector</u>, <u>proper values and vectors</u>.

<u>Quadrature, numerical</u> -- a method for the (approximate) evaluation of a definite integral of a function of a single variable, the term cubature being used for two variables. The most obvious procedure is to obtain some sufficiently close <u>approximation</u> to the integrand f(x), whether by <u>interpolation</u>, by <u>least squares</u>, or otherwise, but in terms that lend themselves readily to analytic integration, and to integrate the result. Any formulas derivable by equal-interval interpolation

can be derived also by the application of difference operators. To do this, let $F(x)$ be the indefinite integral, whence

$$\int_a^b f(x)\, d\, x = F(b) - F(a)$$

and apply the operator $\exp \theta$ to $F(x)$. A somewhat more sophisticated and more general approach is to seek an approximation of the form

$$\int_a^b f(x)\, w(x)\, d\, x = \sum_0^n \mu_i\, f(x_i),$$

where the μ_i are constant coefficients, independent of the function f, and where the expression on the right is to equal the integral on the left exactly whenever $f(x)$ is a polynomial of some maximal degree. When the x_i are specified in advance, this degree can be no greater than n; if they are left free it can be of degree $2n + 1$, and the result is a Gaussian quadrature formula. More general formulas, not based upon polynomial representations, are also possible (see Ho). For special formulas see quadrature formulas under the names Chebyshev, Euler-Maclaurin, Gaussian, three-eighths rule, Simpson rule, trapezoidal rule, Weddle rule; see also remainder formulas.

For solving ordinary differential equations the integrand is of the form $f(x, y)$, where $y = y(x)$ is the dependent variable to be determined, and considered as a function of x,

$$f[x, y(x)] = f(x),$$

it is to be integrated over a range beyond that for which its values are known. Most often it is necessary to express

$$\int_{x_\nu}^{x_{\nu+n}} f(x)\, d\, x$$

in terms of values of f at points x_j, for $j < \nu + n$, in order to obtain an approximate value of $y_{\nu+n}$. Such a formula is called a predictor, or an open quadrature

formula. To simplify the formulas let $\nu = 0$. Given an approximation to $y_{\nu+n}$, one applies a corrector, which is a quadrature formula expressed also in terms of $f_{\nu+n}$.

There is no restriction in taking $\nu = 0$. To obtain a predictor formula, one applies to $F(x_{n-1})$ the operator

$$E - E^{-(n-1)} = e^{\theta} - e^{-(n-1)\theta} = n\,\theta - [(n-1)^2 - 1]\,\theta^2 / 2! + \ldots,$$

and hence to $f(x_{n-1})$ the operator

$$n\,h[1 - (n - 2)\,\theta / 2! + (n^2 - 3n + 3)\,\theta^2 / 3! - \ldots],$$

thereafter substituting

$$\theta = -\ln(1 - \nabla)$$

and retaining as many powers of ∇ in the expansion as may be required. This is the basis for the <u>Adams-Bashford method</u>. To obtain the corrector one applies to $F(x_n)$ the operator $1 - E^{-n}$.

<u>Quotient-difference algorithm</u> -- an algorithm due to Rutishauser and having numerous applications. Let

$$f(z) = s_0 / z + s_1 / z^2 + s_2 / z^3 + \ldots$$

represent a formal power series, and form the table

$s_1^{(0)}$									
	$q_1^{(0)}$								
$s_1^{(1)}$		$d_1^{(0)}$	$e_1^{(0)}$	$s_2^{(0)}$					
	$q_1^{(1)}$				$q_2^{(0)}$				
$s_1^{(2)}$		$d_1^{(1)}$	$e_1^{(1)}$	$s_2^{(1)}$		$d_2^{(0)}$	$e_2^{(0)}$	$s_3^{(0)}$	
	$q_1^{(2)}$				$q_2^{(1)}$				
$s_1^{(3)}$		$d_1^{(2)}$	$e_1^{(2)}$	$s_2^{(2)}$		$d_2^{(1)}$	$e_2^{(1)}$	$s_3^{(1)}$	
	$q_1^{(3)}$								
.	.	.	.	.	.	.	.	.	.

by the following rules:

$$s_1^{(\nu)} = s_\nu,$$

$$q_\sigma^{(\nu)} = s_\sigma^{(\nu+1)} / s_\sigma^{(\nu)}$$

$$d_\sigma^{(\nu)} = q_\sigma^{(\nu+1)} - q_\sigma^{(\nu)}$$

$$e_\sigma^{(\nu)} = d_\sigma^{(\nu)} + e_{\sigma-1}^{(\nu+1)} \quad , \quad e_0^{(\nu)} = 0$$

$$s_{\sigma+1}^{(\nu)} = s_\sigma^{(\nu+1)} \cdot e_\sigma^{(\nu)}.$$

If the entries of a column of q's approaches a limit, this limit is a pole of f(z) (see Bernoulli method). Among other properties of the table, only the following will be mentioned here, that

$$f(z) = s_0 / \overline{z} \; - q_1^{(0)} / \overline{1} \; - e_1^{(0)} / \overline{z} \; - q_2^{(0)} / \overline{1} \; - \ldots$$

is a particular continued fraction expansion of f(z), and others are readily obtainable [NBS].

Radix -- an integer used in a system of numeration whereby all numbers are expressed as sums of powers of it. Thus in the radix r, a number would be written in the form of a sequence of integers

$$\alpha_\nu \, \alpha_{\nu-1} \cdots \alpha_1 \, \alpha_0 \cdot \alpha_{-1} \, \alpha_{-2} \, \alpha_{-3} \cdots$$

with the understanding that $0 \leq \alpha_i \leq r - 1$ for every i, and the interpretation

$$\alpha_\nu \, r_\nu + \ldots + \alpha_1 \, r + \alpha_0 + \alpha_{-1} \, r^{-1} + \alpha_{-2} \, r^{-2} + \ldots.$$

The radix point stands between α_0 and α_{-1}. In the ordinary decimal system the radix is 10; in the binary system it is 2. A digital computer using the radix 3 is said to be under construction in Russia. See bit, binary, decimal.

Radix point -- see radix.

Rank of a matrix -- the number of column (or row) vectors in the smallest set that can be selected from among the columns (or rows) of the matrix that are such that all other columns (or rows) are expressible as linear combinations of them. The number is the same, whether one counts rows or columns. Also it is the order of the largest non-null determinant that can be formed from the rows and columns of the matrix.

Rayleigh-Ritz method -- see Ritz method.

Reciprocal differences -- of a function $f(x)$ whose values are given at points $x_0, x_1, x_2, \ldots$, not necessarily uniformly spaced, are utilized for interpolation by means of rational fractions instead of polynomials. This is appropriate in regions near a singularity. The reciprocal differences

$$\rho(f \mid x_0, \ldots, x_n) = \rho(x_0, \ldots, x_n)$$

are formed by means of a table:

$$\begin{array}{llllll}
x_0 & \rho(x_0) & & & & \\
 & & \rho(x_0, x_1) & & & \\
x_1 & \rho(x_1) & & \rho(x_0, x_1, x_2) & & \\
 & & \rho(x_1, x_2) & & \rho(x_0, x_1, x_2, x_3) & \\
x_2 & \rho(x_2) & & \rho(x_1, x_2, x_3) & & \\
 & & \rho(x_2, x_3) & & \rho(x_1, x_2, x_3, x_4) & \\
x_3 & \rho(x_3) & & \rho(x_2, x_3, x_4) & & \\
 & & \rho(x_3, x_4) & & & \\
x_4 & \rho(x_4) & & & &
\end{array}$$

where

$$\rho(x_i, x_j) = (x_i - x_j) \,/\, [\rho(x_i) - \rho(x_j)],$$

$$\rho(x_i, x_j, x_k) = \rho(x_j) + (x_i - x_k) \,/\, [\rho(x_i, x_j) - \rho(x_j, x_k)],$$

the latter formula being typical. Then the interpolation is made by taking $f(x)$ to be an approximant of the continued fraction

$$f(x) = \rho(x_0) + (x - x_0)/\overline{\rho(x_0, x_1)} + (x - x_1)/\overline{\rho(x_0, x_1, x_2) - \rho(x_0)} + \ldots$$

If the nth approximant is

$$F_n = A_n / B_n,$$

then A_{2p}, A_{2p+1}, B_{2p+1}, B_{2p+2} are of degree p in x. It is not required that the x_i be distinct, and in the extreme case all may coincide. In that event the reciprocal differences are expressed in terms of derivatives, the treatment being analogous to the confluent case of divided differences, and the resulting continued fraction is then analogous to a Taylor series.

Regula falsi -- a method of solving an equation $f(x) = 0$, for a real root, in which if $f(x_0)\, f(x_1) < 0$, and the interval from x_0 to x_1 is known to contain only one real root. One passes the chord from the point $[x_0, f(x_0)]$ to $[x_1, f(x_1)]$, taking x_2, the intersection of the chord with the real axis, as the next approximation. This is

$$x_2 = (x_0\, f_1 - x_1\, f_0) / (f_1 - f_0)$$

where $f_i = f(x_i)$. Repeat, using either x_0 or x_1 with x_2. This is an iterative method which converges more slowly than Newton's, but from the opposite direction, hence if the two are used simultaneously the error is rigorously bounded.

Relative error -- the ratio of the error in an approximation to the true value of the quantity approximated.

Relaxation -- a method of successive approximation for solving a system of equations

$$\varphi_i(x_1, x_2, \ldots, x_n) = 0, \qquad i = 1, 2, \ldots, n.$$

Given any approximate solution $x_i^{(\nu)}$, one evaluates the residuals $\varphi_i(x_j^{(\nu)}) = R_i^{(\nu)}$ and selects the largest. If this is $R_{i'}$, one solves $\varphi_{i'} = 0$ for $x_{i'}^{(\nu+1)}$ in terms of the other $x^{(\nu)}$, recomputes the residuals, and repeats. For machine computation, however, it is generally uneconomical to make the search and the steps are carried

out in a preassigned cyclic order. Hence the term is sometimes extended to include this method. The term was introduced by Southwell, but the method was used by Gauss for solving the normal equations. Some writers use the term for any iterative method (see also matrix inversion).

Remainder -- the difference obtained when an approximation is subtracted from the quantity being approximated. Hence a correction. See remainder formulas.

Remainder formulas -- this term will be applied here to any formula expressing the remainder in approximating a function, or an integral or a derivative of it. A quite general one is the following: Let

$$f(x) = P(x) + R(x)$$

where $P(x)$ is the polynomial of degree n that interpolates $f(x)$ at the points $x_0, x_1, \ldots, x_n$, and let

$$\omega(x) = (x - x_0)\,(x - x_1)\,\ldots\,(x - x_n).$$

Then

$$R(x) = \omega(x) \int_0^1 \int_0^{t_n} \ldots \int_0^{t_1} f^{(n+1)}\,[x + t_0(x_0 - x) + t_1(x_1 - x_0) + \ldots$$

$$+ t_n(x_n - x_{n-1})]\; d\,t_0 \ldots d\,t_n.$$

It is not required that the x_i be distinct, and, in fact, if $x_0 = x_1 = \ldots = x_n$ one has a form of the remainder for Taylor series.

A somewhat simpler form is

$$R(x) = \omega(x)\, f^{(n+1)}\,(\xi)\,/\,(n + 1)!,$$

where, however, of ξ one knows only that it lies on the interval containing all $n + 2$ abscissae: $x, x_0, x_1, \ldots, x_n$.

These same formulas provide remainders for numerical differentiation formulas provided the derivative is to be evaluated at one of the points x_i, since

$$R'(x_i) = \omega'(x_i)\, f^{(n+1)}\,(\xi)\,/\,(n + 1)!.$$

For remainders in the quadrature formulas, refer to these by name.

Remainder theorem -- see algebraic equations.

Ritz method -- for solving a boundary value problem is applicable when the problem can be replaced by an equivalent minimizing problem. The method is to approximate the required function by a member of a family depending upon certain parameters. The problem is then reduced to that of minimizing a certain function of these parameters. The method is closely related to that of Galerkin (see Co).

Root-squaring methods -- methods of obtaining an equation whose roots are the squares of those of a given one. Graeffe's method is the one in common use, although previously Lobačevskiĭ, and still earlier Dandelin, had described other methods and had applied them in the numerical solution of algebraic equations.

Rounding -- the process of dropping the less significant digits in the representation of a number, generally with the rule that a unit should be added in the last place whenever the first discarded digit is 5 or greater in decimal representation, 1 in binary. When this rule is not adopted the process is called truncation.

Rounding error; roundoff -- the error which results from rounding.

Routine -- a sequence of operations carried out by a stored-program computer as specified by a code for the purpose of performing a mathematical or logical computation (cf. program).

Runge-Kutta method -- for the numerical solution of an ordinary differential equation

$$y' = f(x, y),$$

requires the sequential computation of the following quantities:

$$f_\nu = f(x_\nu, y_\nu),$$

$$k_1 = h\, f_\nu, \quad k_2 = h\, f(x_\nu + h/2, \quad y_\nu + k_1/2),$$

$$k_3 = h\, f(x_\nu + h/2, \quad y_\nu + k_2/2), \quad k_4 = h\, f(x_{\nu+1}, y_\nu + k_3),$$

$$y_{\nu+1} = y_\nu + (k_1 + 2k_2 + 2k_3 + k_4)/6.$$

The formulas have an obvious interpretation when y and f are vectors, hence for a system of equations. The method has the advantage of not requiring special measures at the outset as do the methods of Adams, Milne, and others based upon straight interpolation. It has the disadvantage of being slightly more complicated. Analagous simpler schemes, but of lower accuracy, are available (see Co). In this class see the Euler method.

Saddle point method -- for obtaining an asymptotic approximation to a function expressible in the form

$$\int_C e^{zf(\tau)} \, d\tau$$

where $f(\tau) \to -\infty$ at either end of the curve C. If $f(z) = u + i\,v$, the method involves choosing the path C of integration to pass through a saddle point of u, a point that will be among the roots of $f'(\tau) = 0$, and to pass through it in the direction $v = \text{const}$. Points of C in the vicinity of the saddle point then contribute most to the value of the integral, and for large positive z only these points are significant. The method is sometimes called the method of steepest descent since the direction $\sigma = \text{const}$ is that which descends most steeply from the saddle point.

Scalar product -- of two geometric vectors, $\underset{\sim}{x}$ and $\underset{\sim}{y}$, is $|\underset{\sim}{x}|\ |\underset{\sim}{y}| \cos\theta$, where the bars signify the geometric length, and the angle θ is that between the vectors. If these vectors are referred to an orthonormal coordinate system, and if x and y are the column vectors whose elements are the coordinates of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in this system, then

$$x^T y = y^T x = |\underset{\sim}{x}|\ |\underset{\sim}{y}| \cos\theta.$$

In any event, the scalar product is often written $\underset{\sim}{x} \cdot \underset{\sim}{y}$ and hence referred to as the dot product.

Schwarz inequality -- Let φ and ψ be two functions in a space for which an inner product is defined. Then

$$|(\varphi, \psi)|^2 \leq (\varphi, \varphi)\ (\psi, \psi).$$

In particular, if x and y are two vectors (real or complex), then

$$|x^* y|^2 = |y^* x|^2 \leq (x^* x)\ (y^* y)$$

where the asterisk designates the conjugate transpose.

Secular determinant -- for a given matrix A, is

$$\det\ (A - \lambda I),$$

whose zeros are the proper values of the matrix.

Seidel method -- the term sometimes applied to the single-step iteration for solving linear equations. See matrix inversion. Actually the method Seidel (and Gauss) described was a method of relaxation.

Significant figures (or digits) -- those digits that can be considered correct in the approximate representation of a quantity, whether measured or computed, other than zeros to the right of the decimal and to the left of the first non-null digit. Thus in 0.00123, the significant digits are 1, 2, and 3. To write a number in the form 0.001230 is to imply that the final digit is known to be a zero, and that the four digits, 1, 2, 3, and 0 are all significant.

Simplex method -- a method due to George Dantzig for solving the linear programming problem. Let it be required to find that vector x satisfying

$$A\ x = h, \qquad x \geq 0$$

such that $\varphi = \alpha + a^T x$ is minimized, where the matrix A has n rows and $N > n$ columns, and the vectors x and h are of dimensions N and n, respectively. It is assumed that A is of rank n. It is required that initially one find a particular solution x of the inequalities, with $N - n$ of the elements null, and such that the matrix of coefficients of the others is nonsingular.

After a possible renumbering, it can be supposed that the null elements are the last $N - n$. Partition the system with $x^T = (y^T, z^T)$, where y and z are of dimension n and $N - n$, respectively, and write

$$B\,y + C\,z = h, \qquad y = B^{-1}\,h - B^{-1}\,C\,z.$$

At this stage the elements of y are called the basic variables, those of z the nonbasic variables, and the initial solution of the inequalities has $y = B^{-1}\,h \geq 0$ and $z = 0$.

Now φ can be written

$$\varphi = \alpha + b^T y + c^T z = (\alpha + b^T B^{-1} h) + (c^T - b^T B^{-1} C)\,z.$$

If $c^T - b^T B^{-1} C \geq 0$ the problem is already solved. If not, consider one of its negative elements. By allowing the corresponding element of z to become positive, but holding all others at zero, φ can be decreased, but, in general, some element of $y = B^{-1}\,h - B^{-1}\,C\,z$ will eventually vanish. At this point, where the first element of y has vanished, the z which has been allowed to increase is adjoined to the set of basic variables, and the y which has vanished is adjoined to the nonbasic ones, and the process is repeated.

<u>Simpson's rule</u> -- for numerical quadrature:

$$\int_{x_0}^{x_2} f(x)\,d\,x = \frac{h}{3}\,(f_0 + 4f_1 + f_2) + R,$$

where

$$f_i = f(x_0 + i\,h) = f(x_i)$$

$$R = -\,h^5\,f^{IV}\,(\xi)\,/\,90.$$

Hence

$$\int_{x_0}^{x_{2n}} f(x)\,d\,x \doteq \frac{h}{3}\left[f_0 + 4f_1 + 2f_2 + 4f_3 + \ldots + 4f_{2n-1} + f_{2n}\right].$$

Simultaneous equations -- The solution of simultaneous linear equations is discussed under matrix inversion. In case there are more equations than unknowns, it may be that some are redundant, and can be dropped without affecting the theoretical solution. If not, the equations are inconsistent, and only a least squares or other approximation can be used. If there are fewer equations than unknowns the determination cannot be unique, and at most it will be possible to leave some of the unknowns arbitrary, solving for the others in terms of these.

Even with nonlinear equations it is sometimes possible to eliminate some of the unknowns by solving and substituting, although this may not be advisable since the analytic expressions tend to be more complicated. In any case, suppose the system has the form

$$\varphi_i(\xi_1, \ldots, \xi_n) = 0, \quad i = 1, 2, \ldots, n,$$

assume no further reduction is possible or desirable, and suppose that a solution $\alpha_1, \alpha_2, \ldots, \alpha_n$ actually exists. If

$$\Phi = \sum_1^n \varphi_i^2,$$

then $\Phi(\xi_1, \ldots, \xi_n) \geq 0$ and $\Phi(\alpha_1, \ldots, \alpha_n) = 0$. Hence Φ is minimized for $\xi_i = \alpha_i$. Therefore one may use the method of steepest descent as one method of solving the system, provided the necessary derivatives exist.

Another method is a generalization of Newton's method. Suppose $\xi_i^{(0)}$ is a set of numbers sufficiently close to α_i. Then, by expanding in Taylor's series,

$$\varphi_i(\alpha_1, \ldots, \alpha_n) = 0 = \varphi_{i,0} + \sum_j (\alpha_j - \xi_j^{(0)}) \, \partial \varphi_{i,0} / \partial \xi_j + \ldots,$$

where the subscript 0 signifies evaluation for $\xi_i = \xi_i^{(0)}$. If the differences $\alpha_j - \xi_j^{(0)}$ are small enough and the terms not written (quadratic and higher) are negligible, and if, further, the matrix of first partial derivatives

$$(\partial \varphi_i / \partial \xi_j)$$

is nonsingular throughout a sufficiently large neighborhood of the solution α_j, then the linearized equations can be solved and the result will be approximately equal to the differences $\alpha_j - \xi_j^{(0)}$. By adding these to the $\xi_j^{(0)}$ one obtains an improved set, and the process can be continued as many times as seems necessary.

Single-step iteration -- see matrix inversion.

Singular values -- of a matrix A are the non-negative square roots of the proper values of A* A.

Smoothing -- the replacement of a curve, or of a sequence of points by another that is in some sense more regular, and yet whose ordinates, for any abscissa, are changed as little as possible. The irregularities in a sequence of points may be due to errors in measurement. If theory requires the theoretically correct points to lie on a given curve, one may apply some method of curve fitting, possibly least squares. If not, one may select arbitrarily a simple function, possibly a polynomial, and fit it by least squares. If the purpose is merely to obtain a smooth graph, this may be drawn visually. Somewhat more sophisticated is to take, say, 5 consecutive points, fit a parabola, and replace the middle point by the one on the parabola. The next parabola requires four of these points and one new one.

Spectral norm -- This term is sometimes used for the spectral radius, but this function does not possess the norm properties. Also, and more appropriately, applied to the function $\|A\|$ defined by

$$\|A\|^2 = \rho(A^* A),$$

which is a norm (see spectral radius).

Spectral radius of a matrix A -- $\rho(A) = \max_i |\lambda_i(A)|$ for all proper values of A. See norms; spectral norm.

Stability -- see numerical stability.

Star -- the set of coefficients associated with any particular approximation to derivatives by differences used in the numerical solution of partial differential equations.

Steepest descent -- a term sometimes applied to the saddle point method. However the term is also applied to a method of solving simultaneous equations and of minimizing (or maximizing) a function. Given a function of several variables

$$\Phi(\xi_1, \xi_2, \ldots, \xi_n)$$

to be minimized, form grad Φ, the vector whose elements are $\partial \Phi / \partial \xi_i = \gamma_i$. The function

$$\varphi(\lambda) = \Phi(\xi_1 - \lambda \gamma_1, \ldots, \xi_n - \lambda \gamma_n)$$

is a function of only a single variable. Let $\xi_1^{(0)}, \ldots, \xi_n^{(0)}$ represent a point sufficiently close to the minimum, let $\gamma_i^{(0)}$ represent the value of γ_i at that point; substitute $\xi_i^{(0)} - \lambda \gamma_i^{(0)}$ in Φ and minimize the resulting φ as a function of λ. If this minimum occurs at $\lambda^{(0)}$, then $\xi_i^{(1)} = \xi_i^{(0)} - \lambda^{(0)} \gamma_i^{(0)}$ will be, in general, a closer approximation to the minimum.

Stirling formula -- if n is a large number

$$\ell n\, n! = \left(n + \frac{1}{2}\right) \ell n\, n - n + \frac{1}{2} \ell n\, 2\pi + \frac{1}{12n} - O(1/n^3),$$

where the last term is of the order of $1/n^3$. Often approximated by:

$\ell n\, n! = n\, \ell n\, n - n.$

Stirling interpolation formula -- in the notation of difference operators

$$E^u = 1 + u\,\mu\,\delta + \frac{u^2}{2!}\delta^2 + \frac{u(u^2-1^2)}{3!}\mu\,\delta^3 + \frac{u^2(u^2-1^2)}{4!}\delta^4 + \frac{u(u^2-1^2)(u^2-2^2)}{5!}\mu\,\delta^5 + \ldots$$

Sterling numbers -- the coefficients $S_\nu^{(n)}$ in the expansion (see difference operators)

$$u^{(n)} = S_n^{(n)} u^n + S_{n-1}^{(n)} u^{n-1} + \ldots + S_1^{(n)} u.$$

For $\nu > n$, $S_\nu^{(n)} = 0$. These are also called Stirling numbers of the first kind, those of the second kind being the coefficients $\mathfrak{S}_r^{(n)}$ of the reverse expansion

$$u^n = \mathfrak{S}_n^{(n)}\, u^{(n)} + \mathfrak{S}_{n-1}^{(n)}\, n^{(n-1)} + \ldots + \mathfrak{S}_1^{(n)}\, u^{(1)}.$$

It follows from these definitions that

$$S_\nu^{(n)} = \left.\frac{d^\nu\, u^{(n)} / d\, u^\nu}{\nu!}\right|^{u=0}, \qquad \mathfrak{S}_\nu^{(n)} = \left.\frac{\Delta^\nu\, u^n}{\nu!}\right|^{u=0}.$$

Their importance is due in part to their occurrence in the following <u>differentiation formulas</u>:

$$\theta^\nu = S_\nu^{(\nu)}\, \Delta^\nu + \frac{S_\nu^{(\nu+1)}\, \Delta^{\nu+1}}{\nu + 1} + \frac{S_\nu^{(\nu+2)}\, \Delta^{\nu+2}}{(\nu+1)(\nu+2)} + \ldots$$

$$\Delta^\nu = \mathfrak{S}_\nu^{(\nu)}\, \theta^\nu + \frac{\mathfrak{S}_\nu^{(\nu+1)}\, \theta^{\nu+1}}{\nu + 1} + \frac{\mathfrak{S}_\nu^{(\nu+2)}\, \theta^{\nu+2}}{(\nu+1)(\nu+2)} + \ldots$$

<u>Storage unit</u> -- that part of a <u>stored-program computer</u> which stores the <u>code</u> or the numerical data or both. In most current machines the storage unit is divided into cells, usually of between 30 and 50 binary elements, these cells are ordered and the ordinal number of the cell is its address. Generally, too, a given cell can be used to store either data or commands, a command including an operational code symbol and one or more address symbols for locating the operands. Operation can begin with any cell, the contents being interpreted as one or more commands, and will usually proceed in sequence until a transfer or jump command interrupts, although in some machines every command specifies also the location of the next command to be executed.

<u>Stored-program computer</u> -- a computer that is capable of storing instructions as well as data and of carrying out a sequence of operations as prescribed in advance. The modern high-speed general-purpose computers are <u>digital computers</u> of this type. Any such machine will possess a basic repertory of certain arithmetic and

logical operations, each designated by a certain symbol, basically a binary sequence. A sequence of such symbols constitutes a code which, when introduced into the storage unit along with the necessary data will direct the machine accordingly.

Störmer method -- for the numerical solution of an ordinary differential equation of second order;

$$y'' = f(x, y, y'),$$

uses the predictor

$$y_{\nu+1} = 2y_\nu - y_{\nu-1} + h^2 \left(f_\nu + \frac{1}{12}\nabla^2 f_\nu + \frac{1}{12}\nabla^3 f_\nu + \frac{19}{240}\nabla^4 f_\nu + \frac{3}{40}\nabla^5 f_\nu + \ldots\right).$$

It is useful in particular when y' does not occur explicitly in f. The equations apply also when y and f are vectors. See difference operators for the definition of ∇. (See Co). *

Sturm theorem -- for locating roots of an algebraic equation $P(x) = 0$. Taking $P = P_0$, $P' = P_1$, apply the Euclidean algorithm for the highest common factor of P_0 and P_1. In the Sturm sequence $P_0, P_1, P_2, \ldots$ evaluated at any r, let V_r represent the number of variations in sign. Then the equation has exactly V_r distinct roots exceeding r. While the application of Budan's theorem is simpler, it provides less information in general. Moreover, if P_i is not a constant, P_i being the last term in the sequence, all multiple roots satisfy $P_i = 0$ (see algebraic equations).

Subroutine -- a portion of a routine for a stored-program computer, e.g., a part which evaluates a trigonometric or other standard function, which has been so coded that it can be made an integral part of any routine that calls for this particular computation.

* Sturm Sequence -- a sequence of polynomials associated with a given polynomial possessing the properties required for applying the Sturm theorem, that no consecutive pair vanish simultaneously, and that when one vanishes the adjacent ones are opposite in sign.

Summation of series -- In general, methods of approximate computation are based upon the use of infinite series and sequences whose limits are the desired quantities, and in practice one must truncate the series or select a particular term in the sequence. It is advantageous, therefore, that convergence be as rapid as possible. In some instances a series may not converge and yet may be regarded as defining the required quantity in some sense. Thus the series $1 + x + x^2 + x^3 + \ldots$ fails to converge when $|x| > 1$; on the other hand it is the formal development of $(1 - x)^{-1}$ and may be considered to define this quantity whether or not the series converges. Moreover, if one applies the delta-square process to the sequence of partial sums of this series one obtains precisely $(1 - x)^{-1}$. Thus the process, in this instance, transforms the sequence of partial sums into a new sequence of which every term is the value of the quantity formally defined by the series. It can be regarded as a device for accelerating convergence, when convergence occurs, or of converting from divergence to convergence otherwise, in either event a method for summing the series.

Linear methods of summation have been most exhaustively studied, and the Euler transformation is one of the simplest and most useful. The delta-square process is among the simpler and more powerful of known nonlinear methods. The elements u_i of an infinite sequence can always be regarded as the partial sums of a series, hence one can direct attention to a series or to a sequence, according to convenience. The terms u_n of a sequence can always be regarded as functional values $u(x)$ of a function for which $x = n^{-1}$. Hence finding the limit of a sequence of terms u_n can be regarded as equivalent to the problem of extrapolating the function $u(x)$ to $x = 0$, when functional values are known for $x = n^{-1}$. From this point of view one can use interpolation polynomials, continued fractions (e.g., formed with reciprocal differences) or any other convenient method of approximation.

Symmetric function -- of n variables $x_1, \ldots, x_n$, any function $f(x_1, x_2, \ldots, x_n)$ not affected by an interchange of any x_i and x_j. If these are roots of an algebraic equation

$$x^n - c_1 x^{n-1} + c_2 x^{n-2} - \ldots \pm c_n = 0,$$

then

$$c_1 = \sum x_i, \quad c_2 = \sum x_i x_j, \quad c_3 = \sum x_i x_j x_k, \ \ldots$$

are the elementary symmetric functions, where the summations are extended over all distinct products of distinct factors, the x_i being themselves considered independently varying. Any rational symmetric function of the x_i is a rational function of the c_i.

Synthetic division -- To divide the polynomial

$$P(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots$$

by $x - r$, write

$$\begin{array}{ccccc|l} a_0 & a_1 & a_2 & \ldots & a_n & r \\ & a_0 r & b_1 r & & b_{n-1} r & \\ \hline a_0 & b_1 & b_2 & & R & \end{array}$$

where each number below the line is the sum of those above. Then

$$Q(x) = a_0 x^{n-1} + b_1 x^{n-2} + b_2 x^{n-3} + \ldots,$$

where

$$P(x) = (x - r)\, Q(x) + R.$$

Taylor series -- for a function $f(x)$ of a single variable has the form

$$f(x) = f(a) + (x - a) f'(a) + \ldots + (x - a)^n f^{(n)}(a) / n! + R_n ,$$

assuming the derivatives to exist, where R_n is the remainder. There are many different ways of expressing this, among them being

$$R_n = \frac{(x - a)^{n+1}}{n!} \int_0^1 (1 - t)^n f^{(n+1)} [a + t (x - a)] \, d t;$$

$$R_n = \frac{(x - a)^{n+1}}{(n + 1)!} f^{(n+1)} (\xi), \qquad a \leq \xi \leq x.$$

If $f(x_1, x_2, \ldots, x_n)$ is a function of n variables the expansion has the form

$$f(x_1, x_2, \ldots, x_n) = f(x_1 - a_1, \ldots, x_n - a_n) + \sum (x_i - a_i) f_i(a_1, \ldots, a_n)$$
$$+ \frac{1}{2!} \sum \sum (x_i - a_i) (x_j - a_j) f_{ij} (a_1, \ldots, a_n) + \ldots$$

where the subscripts on f signify partial differentiation.

Tchebychev -- see Chebyshev.

Three-eighths rule -- the quadrature formula

$$\int_{x_0}^{x_3} f(x) \, d x = 3h(f_0 + 3f_1 + 3f_2 + f_3) / 8,$$

where

$$f_i = f(x_i) = f(x_0 + i h).$$

The remainder is

$$R = - 3h^5 f^{IV} (\xi) / 80,$$

hence the formula is in general somewhat less exact than the simpler Simpson's rule.

Trace of a matrix A -- $\quad \text{trace } (A) = \sum \alpha_{ii}$

where $A = (\alpha_{ij})$.

Transfer of control -- in the usual stored-program computers commands are executed in the sequence in which they occur in storage unless a command of a specific type (a transfer command) interrupts the sequence and causes the machine to look elsewhere for its next instruction. The command may be conditional, to be executed only in case a certain condition is satisfied, or unconditional.

Trapezoidal rule -- for numerical quadrature:

$$\int_a^b f(x)\, d\, x = (b - a)\, [f(a) + f(b)] / 2 + R,$$

$$R = - h^3 f''(\xi) / 12, \qquad a \le \xi \le b, \; h = b - a$$

Triangular matrix -- one in which all elements are ~~null~~ zero above the diagonal (a lower triangular matrix), or else below the diagonal (an upper triangular matrix). It is a unit upper or lower triangular matrix ~~when also all diagonal elements are units~~ when also every diagonal element $= 1$. The product of two triangular matrices of the same type is again of that type, i.e., (unit) upper or lower; and the reciprocal of a triangular matrix (when it exists) is also of the same type. A triangular matrix is properly triangular if the diagonal is null.

Trigonometric interpolation -- Formulas analogous to the Lagrange formula for polynomial interpolation can be obtained. They all have the form

$$f(x) \doteq \sum f_i\, T_i(x) / T_i(x_i).$$

Possible choices of $T_i(x)$ are the following:

$$T_i(x) = \prod_{j \neq i} \sin\, (x - x_j) / 2 \;;$$

$$T_i(x) = \prod_{j \neq i} (\cos x - \cos x_j) \;;$$

$$T_i(x) = \sin x \prod_{j \neq i} (\cos x - \cos x_j) \;;$$

$$T_i(x) = \prod_{j \neq i} (\sin x - \sin x_j) \;.$$

Trigonometric Series -- See Fourier Series.

Triple-diagonal matrix -- a matrix $A = (\alpha_{ij})$ for which $\alpha_{ij} = 0$ when $|i - j| > 1$. Thus nonnull elements occur only along, just above, or just below the diagonal. These occur in the theory of continued fractions. Some methods for computing proper values and vectors of a matrix aim first at transforming the matrix to triple-diagonal form.

Truncation -- approximating the sum of an infinite series by the sum of a finite number of terms in that series. As applied to the dropping of digits in a number, the term contrasts with rounding.

Truncation error -- strictly the error due to dropping all but a finite number of terms from a possibly infinite series, but often applied also to the error due to representing the limit of a sequence by one of its terms, or to representing a function by an interpolation polynomial. See error.

Unitary matrix -- see matrix.

Vector -- a matrix of n rows and a single column. Geometrically or physically, it can represent directional magnitudes such as displacements, velocities, accelerations, forces, etc.

Volterra equation -- the integral equation of the form

$$\varphi(x) = f(x) - \lambda \int_{x_0}^{x} K(x, y)\, \varphi(y)\, d\, y$$

in which the upper limit of integration is the variable x.

Wave equation -- the hyperbolic partial differential equation of the form

$$\partial^2 u / \partial x^2 = \kappa^2\, \partial^2 u / \partial t^2.$$

Weddle rule -- for numerical quadrature,

$$\int_{x_0}^{x_6} f(x)\, d\, x = 3(f_0 + 5f_1 + f_2 + 6f_3 + f_4 + 5f_5 + f_6) + R,$$

$$R = - h^7 \, [10 \, f^{(VI)} \, (\xi_1) + 9h^2 \, f^{(VIII)} \, (\xi_2)] \, / \, 1400.$$

Word -- in the terminology of digital computers, a set of characters occupying one storage cell and treated as a unit by the machine, e.g., in being transferred to and from registers.

SELECTED REFERENCES

[with abbreviations used when specific reference is made in the text]

N. I. Achieser (Ahiezer) (1956): Theory of approximations (tr.), N. Y. Ungar, x + 307 p. - Ah

Friedrich L. Bauer (1954): Beiträge zur Entwicklung numerischer Verfahren für programmgestenerte Rechenanlagen. I. Quadratisch Konvergente Durchführung der Bernoulli-Jacobischen Methoden zur Nullstellenbestimmung von Polynomen, Akad. Wiss München. S:tz. Math.-Natur. Kl. 275-303.

Friedrich L. Bauer (1956): Beiträge zur Entwicklung numerischer Verfahren für programmgesteuerte Rechenanlagen II. Direkte Faktorisierung eines Polynoms, Bay. Ak. Wiss. Math.-Nat. Kl. Sitz. 163-203.

Lothar Collatz (1955): Numerische Behandlung von Differentialgleichungen. Zweite neubearbeitete Auflage. Springer, xv + 526 p. - Co

W. Edwards Deming (1938): Statistical adjustment of data, New York: John Wiley & Sons, Inc., pp. x + 261. - De

Wallace Givens (1954): Numerical computation of the characteristic values of a real symmetric matrix, ORNL 1574.

Cecil Hastings, Jr., Jeanne T. Hayward, and James P. Wong, Jr. (1955): Approximations for digital computers, Princeton Univ. Press, 201 pp. - Ha

Magnus R. Hestenes and Eduard Stiefel (1950): Methods of conjugate gradients for solving linear systems., JRNBS 49:409-36.

F. B. Hildebrand (1956): Introduction to numerical analysis, New York, Toronto, London, McGraw-Hill, x + 511 pp. - Hi

Alston S. Householder (1953): Principles of numerical analysis, N. Y. McGraw-Hill, x + 274 pp. - Ho

A. S. Householder, G. E. Forsythe, and H. H. Germond (eds.) (1951): Monte Carlo method, NBS Appl. Math. Series 12, vii + 42 pp. - MC

Zdenek Kopal (1955): Numerical analysis, New York, John Wiley & Sons, xiv + 556 pp. - Ko

Kaiser S. Kunz (1957): Numerical analysis, N. Y.-Toronto-London, McGraw-Hill, xv + 381 pp. - Ku

Cornelius Lanczos (1956): Applied analysis, Englewood Cliffs, N. J., Prentice Hall, 539 pp. - La

H. A. Meyer (ed) (1956): Symposium on Monte Carlo methods, held at University of Florida, 1954., New York, John Wiley & Sons, xvi + 382 pp. - sym

Philip M. Morse and Herman Feshbach (1953): Methods of theoretical physics, New York, McGraw-Hill, xxii + 1978 pp. - M & F

National Bureau of Standards (1958): Further contributions to the solution of simultaneous linear equations and the determination of eigenvalues, NBS Appl. Math. Series 49 - NBS

National Physical Laboratory (1957): Modern computing methods, Notes on applied science No. 16, London, Her Majesty's Stationery Office, vi + 129 pp. - NPL

Heinz Rutishauser (1957): Der Quotienten-Differenzen-Algorithmus, Basel/Stuttgart, Birkhäuser Verlag, 74 pp.

Daniel Shanks (1955): Non-linear transformations of divergent and slowly convergent sequences, JMP 34:1-42 pp.

E. T. Whittaker and G. Robinson (1940): The calculus of observations. A treatise on numerical mathematics, London and Glasgow: Blackie & Sons, Ltd., [redacted] ed. pp. xvi + 395 - W & R
3rd

Kaj Leo Nielsen (1956): Methods in numerical analysis, New York, Macmillan, 328 pp. - NI

Niels Erik Nörlund (1924): Vorlesungen über Differenzenrechnung, Berlin, Springer, ix + 551 pp. - NO

ORNL-2704
Physics and Mathematics
TID-4500 (14th ed.)

INTERNAL DISTRIBUTION

1. C. E. Center
2. A. M. Weinberg
3. L. B. Emlet (K-25)
4. J. P. Murray (Y-12)
5. J. A. Swartout
6. E. H. Taylor
7. E. D. Shipley
8. A. S. Householder
9. C. P. Keim
10. W. H. Jordan
11. S. C. Lind
12. F. L. Culler
13. R. S. Livingston
14. A. H. Snell
15. A. Hollaender
16. M. T. Kelley
17. K. Z. Morgan
18. T. A. Lincoln
19. C. S. Harrill
20. H. E. Seagren
21. C. E. Winters
22. D. Phillips
23. M. L. Nelson
24. J. A. Lane
25. R. A. Charpie
26. M. J. Skinner
27. R. R. Dickison
28. P. M. Reyling
29. G. H. Stakes
30. E. Leach
31. C. A. Preskitt
32. N. Alexander
33. G. J. Atta
34. S. E. Atta
35. H. P. Carter
36. E. L. Cooper
37. A. H. Culkowski
38. N. M. Dismuke
39. A. C. Downing, Jr.
40. Ky Fan
41. M. Feliciano
42. C. T. Fike
43. D. A. Gardiner
44. A. A. Grau
45. J. Harrison
46. M. A. Kastenbaum
47. A. W. Kimball
48. E. C. Long, Jr.
49. D. J. Wehe
50. J. P. Kelly (K-25)
51. J. E. Rowe (K-25)
52. Biology Library
53. Health Physics Library
54-55. Central Research Library
56. Reactor Experimental Engineering Library
57-276. Laboratory Records Department
277. Laboratory Records, ORNL R.C.
278. ORNL - Y-12 Technical Library, Document Reference Section

EXTERNAL DISTRIBUTION

279. Division of Research and Development, AEC, ORO
280-815. Given distribution as shown in TID-4500 (14th ed.) under Physics and Mathematics category